**The Dissertation Committee for Alexander Crofts Certifies that this is the approved version of the following Dissertation:**

# Bacterial Pathogen Adaptation During Human Infections

**Committee:**

Bryan W. Davies, Supervisor

Michael Trent, Co-Supervisor

Shelley M. Payne

Howard Ochman

Maria A. Croyle

# Bacterial Pathogen Adaptation During Human Infections

**by**

**Alexander Crofts**

**Dissertation**

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

**Doctor of Philosophy**

**The University of Texas at Austin**

**August 2018**

# Dedication

To my wife, Lindsay.

## Acknowledgements

First, I thank my wonderful parents for supporting me in science from elementary school through today. To my brother, whose intellect and accomplishments are inspirational – you're the original Crofts Ph.D. I also thank my sister (-in-law), whose invaluable insight and endless motivation got me into, through, and beyond graduate school.

I became qualified for graduate school thanks, in part, to people who did not check my undergraduate GPA before hiring me. The experiences I gained with Drs. Isaac Cann, Ethan Chen, Taekjip Ha, Sinan Arslan, Juliane Bubeck-Wardenburg, Rebecca Spokony, Matthew Slattery, and Kevin White got me to where I am today.

Although, my Jedi Master Ph.D. mentor, Stephen Trent, was sure check my transcripts and let me know my chemistry grades were terrible. Fortunately, he allowed me to join his lab and also graciously allowed me to remain in Austin when he moved to Georgia. Stephen, you were a better mentor than I could have ever hoped for, and I hope I wasn't too squirrely. I'm also sorry *yodA* didn't work out – the powerpoints would have been amazing. After I joined the lab, the support I received from Jeremy Henderson, Emily Nowicki, Brittany Needham, and Carmencita, was second to none. Later on, my step-lab-dad Bryan Davies, Ashley Tucker, Joseph Boll, and the irreplaceable Simone Giovanetti were among other wonderful labmates who I could not have survived without. My fantastic collaborator, Dr. Patricia Guerry, her team, and the brave people who volunteered to be infected enabled the entirety of this work.

Finally, my amazing wife Lindsay, who continues to leave her careers behind so I can pursue mine. You're the best high school chemistry partner I could have ever dreamed of!

# Abstract

## Bacterial Pathogen Adaptation During Human Infections

Alexander Crofts, Ph.D.

The University of Texas at Austin, 2018

Supervisors: Bryan W. Davies, Michael Trent

Like all organisms on earth, bacteria must adapt to changes in their environment to survive. Thus, discovering bacterial adaptations reveals the tools bacteria use to be successful. Identifying how pathogenic bacteria adapt during infections can consequently identify the tools bacteria use to cause disease, and therapy design can then consider inhibiting these tools to treat or prevent infections. Here, the ways in which two worldwide human intestinal pathogens, *Campylobacter jejuni* and Enterotoxigenic *E. coli* (ETEC), adapt to the human host during infections are explored. Bacteria were studied directly in infected samples from controlled human infection models. In *C. jejuni*, genetic adaptations that were selected for during acute and persistent human infections identified the role of a previously uncharacterized flagellar modification gene during persistence. In ETEC, the bacteria's ability to sense oxygen was linked to global virulence gene expression in human infection samples as well as biofilm formation. As environmental ETEC biofilms are associated with seasonal ETEC epidemics, oxygen sensing likely contributes to human infection inside and outside of the host. Together, these data demonstrate the scope of pathogen adaptation during infections, identified new targetable virulence factors, and can thus aid the design of new therapies.

# Table of Contents

# List of Tables

# List of Figures

# List of Illustrations

# Chapter 1: Introduction

## 1.1 Overview

Bacterial intestinal infections are an unparalleled burden on humanity. Billions of days of human productivity are lost annually due to symptomatic intestinal disease[1]. The most prominent symptom of bacterial intestinal infections is diarrhea, and profuse diarrhea can lead to dehydration and death, particularly in children of low resource areas[1–3]. This dissertation focuses on two major global intestinal pathogens, *Campylobacter jejuni* and Enterotoxigenic *Escherichia coli* (ETEC), which were first identified as causes of infectious diarrhea within a year of each other in the early 1970s[4,5]. Chapter 1 provides background information about these Gram-negative bacteria and our continued need to better understand how they cause human disease. One approach to understanding disease development is to study the bacterial pathogens directly in infected human samples[6–9], and this is also discussed in Chapter 1. In particular, the utility of characterizing bacterial adaptation to the host environment is highlighted.

Rare samples collected from controlled human infection models of *C. jejuni* and ETEC infections were used throughout this work. Chapter 2 demonstrates how *C. jejuni* adapts to the human host by using high-throughput nucleic acid sequencing for genomic and transcriptomic analyses. These data revealed new insight into how *C. jejuni* causes human disease along with new therapeutic targets. Chapter 3 demonstrates how ETEC senses a hallmark of the intestinal environment to regulate virulence factor expression during human infections. Together, these data are discussed in Chapter 4 for their contribution to understanding bacterial adaptation during human infection. Finally, the methods used throughout this work, along with cited material, are provided.

## 1.2 CAMPYLOBACTER JEJUNI: A ZOONOTIC PATHOGEN OF THE INTESTINAL TRACT

### 1.2.1 Impact and Sources of *C. jejuni* Infections

*Campylobacter jejuni* is spiral-shaped, Gram-negative bacteria that was isolated from human feces and identified to be a cause of human intestinal infection in 1972[5]. It is currently the leading cause of bacterial gastroenteritis worldwide and causes approximately 1 million infections in North America annually[10].

Symptoms of *C. jejuni* infection vary widely between patients - from asymptomatic colonization to dysentery and bloody diarrhea. Fever and severe abdominal cramps are common. A standard presentation of *C. jejuni* infection can start with fever, cramps, and headache before bouts of diarrhea begin that may continue for 5-7 days. Typically, infections are self-limiting and do not require antibiotic treatment. Although, approximately 15% of gastroenteritis hospitalizations are due to *C. jejuni* infection[10].

*C. jejuni* infections are almost exclusively caused by ingestion of contaminated food or water. *C. jejuni* is a commensal bacterium in the intestines of wild and domesticated birds - most notably in chickens – and ingestion of contaminated poultry products is a major source of *C. jejuni* infections[2,11]. Other livestock, including cattle and sheep, can also be seemingly asymptomatically colonized by *C. jejuni*. Therefore, contaminated milk can also be a source of infection[12]. In high-resource areas, ingestion of contaminated and inappropriately prepared poultry meat is the leading cause of *C. jejuni* infection[10]. In low-resource areas, contaminated milk, recreational waters, or exposure to bird/animal droppings in the environment are major sources of infection[2,13].

Observation of a spiral shaped bacterium associated with *Vibrio cholerae*-like diarrhea in children were reported in the 19th century, however the organism could not be cultured[14]. Similar observations were attributed to "*Vibrio*-like" bacteria in later cases,

until the bacteria was finally cultured from the feces of a human patient in 1972[5]. The patient had a fever and diarrhea and it was presumed an intestinal pathogen had become septicemic. Filtering the patient's feces and incubating the filtrate on solid medium in low-oxygen conditions enabled *C. jejuni* to be cultivable, and it was identified in both the diarrhea and the blood stream of the patient.

In the laboratory, *C. jejuni* is fastidious, with a sensitivity to refrigeration and freezing, atmospheric concentrations of oxygen, and is asaccharolytic, so standard carbohydrates energy sources for use *in vitro* are not effective[15,16]. This likely explains why previous culturing attempts had failed. This fastidiousness is seemingly at odds with the widespread incidence of *C. jejuni* infections and understanding this discrepancy is an active area of research[17].

Infrequent yet serious extra-intestinal complications can also arise. *C. jejuni* intestinal infections can lead to septicemia, as *C. jejuni* can be isolated from blood and can pose a risk to the health of a fetus in pregnant women[18]. Reactive arthritis of autoimmune origin can also occur which presents as painful and stiff joints which limit patient mobility[19]. Sometimes this presents before diarrhea occurs and the exact C. jejuni factors that induce this arthritis are yet to be determined.

*C. jejuni* infection also can also induce another autoimmune disorder, called Guillain-Barré Syndrome (GBS)[20], which is a life-threatening paralysis that requires immediate medical intervention. GBS results from anti-*C. jejuni* antibodies that cross-react with surface exposed sugars found on nerve cells[21]. The resulting autoimmune response degrades neurons causing a progressive paralysis which can quickly lead to diaphragm dysfunction and suffocation. Immunosuppressants can halt the response and eventually allow for recovery[22]. The cross-reactive antibodies are generated by immune recognition of sugars on *C. jeuni* lipooligsaccharide that structurally mimic human

gangliosides[23]. These antibodies therefore recognize both *C. jejuni* and neuron surface structures. A naturally occurring *C. jejuni* strain, CG8421, was isolated from a patient in Thailand which is missing the genetic loci that encode for the production of the cross-reacting sugars[24]. CG8421 was therefore approved for use in controlled human infection models of infection as the risk of potentially lethal GBS was eliminated. Humans infected with CG8421 produce a typical range of symptomatic disease, ranging from lose stools to severe bloody diarrhea with high fever and abdominal cramping[24].

Besides the high incidence of infection in high-resource areas, children in low-resource areas are at risk of lethal or permanently debilitating infections[13]. Recent exhaustive global surveys have associated Campylobacter infections during the first year of life with stunted growth periods[2]. Growth stunting is thought to occur due to malnutrition during infection and is associated with lifelong physical and cognitive deficits. Consequently, what is sometimes considered to be a nuisance bout of diarrhea in high-resource areas can be a devastating, life-changing event in low-resource areas.

### 1.2.2 *C. jejuni* Colonization Factors

The first *C. jejui* genome sequence was surprising for a number of reasons - most notably for revealing a complete lack of pathogen-defining toxins[25]. The genome was also surprisingly small (about 1/3 the size of the ETEC genome) and harbored seemingly no unique or notable secreted virulence factors. For instance, there are no homologs to enterotoxins, cholera toxin, or shiga toxin in the genome. Another surprising feature was over 25 phase variable regions where homopolynucleotide C/G tracts were found throughout the genome. These phase variable regions are prone to insertion/deletion mutations and consequently introduces astounding genetic and phenotypic heterogeneity in a *C. jejuni* population[26].

Overall, there is seemingly minimal mechanistic understanding of *C. jejuni* human pathogenesis when compared to other intestinal pathogens that have clearly defined essential virulence factors. It is believed that despite our knowledge of the genome, there are likely conserved, defining virulence factors yet to be discovered in *C. jejuni*[27]. That notwithstanding, prominently accepted *C. jejuni* colonization factors and phase variation's contribution to disease are covered in the remaining sections of this chapter.

#### *1.2.2.1 Motility, Adherence, and invasion*

Unlike other intestinal pathogens that rely on characteristic and essential adhesins and secreted toxins for colonization, major factors dictating *C. jejuni* colonization are considered to be basic observations of motility, adherence, and invasion. It is an invasive pathogen, although mechanisms of adhesion and invasion into host cells *in vivo* are not well understood. After ingestion *C. jejuni* uses flagellar-based motility to enter the mucus layer that protects the small intestine epithelium, and consequently adheres/associates to the host epithelium[28]. After colonizing the small intestine, *C. jejuni* can also colonize the colon which is likely to origin of blood when bloody diarrhea symptoms present[29].

*In vitro*, *C. jejuni* can be seen adhering to and invading epithelial cells. JlpA[30,31] and CadF[32] are two *C. jejuni* surface exposed proteins have been shown to bind to host surface/extracellular proteins (Hsp90 and fibronectic, respectively) to mediate adhesion *in vitro*. After adhesion C. jejuni can invade host cells, although mechanisms of invasion are murky. A group of proteins called Campylobacter invasion antigens (Cia) have been identified that enhance the rate of *C. jejuni* invasion[33], and CiaB is internalized into host cells via unknown mechanisms[33].

Some strains are more invasive than others, indicating a role for distinct genetic loci beckoning the discovery of new invasins. A gene called *cipA* was identified in unbiased screens to be associated with strains that are more invasive and with strains that caused more severe human disease[34]. Although until this work presented in Chapter 2, the biochemical function of CipA was unknown.

Interestingly, *C. jejuni* lacks type III secretion mechanisms that other invasive bacteria use to secrete invasins. However, it was reported that *C. jejuni* surprisingly uses the flagellar apparatus to secrete proteins[35], possibly in lieu of a type III secretin system. These secreted proteins include the Cia proteins along with other non-flagellar proteins[36]. For instance, the *C. jejuni* specific protein FspA (flagella secreted protein A) is secreted through the flagellar filament, and when added exogenously onto epithelial cells *in vitro*, induces host cell apoptosis. However, the role of this secretion mechanism or these secreted proteins in mammalian infections is unresolved.

#### *1.2.2.2 Cytolethal Distending Toxin*

The heterotrimeric AB toxin, cytolethal distending toxin (CDT), is found in multiple pathogenic *Proteobacteria*, including *Shigella dysenteria*, *E. coli*, *Salmonella enterica*, and *C. jejuni*[37]. The CdtB subunit of the CdtABC holotoxin uses DNAse activity to arrest eukaryotic cells in G2/M of the cell cycle[38]. This inhibition results in large, distended cells which give the toxin its name.

Because of work done on CDT in the context of other pathogens, CDT is inadvertently the most well characterized toxin in *C. jejuni*[39]. However, the role of CDT in symptomatic human disease or chicken colonization is unclear. Notably, symptomatic human disease can be caused by CDT-negative *C. jejuni* strains[40], and its role in the native host is unclear as chickens do not make antibodies against CDT[40,41]. Further, other

pathogens that harbor CDT do not necessarily use it as an essential colonization factor[42], suggesting a limited role in *C. jejuni* as well.

#### *1.2.2.3* Iron Acquisition

Iron is a vital cofactor for all bacteria, and host organisms have numerous mechanisms for sequestering iron as a form of nutritional immunity[43,44]. To increase the efficiency of iron acquisition, many bacteria secrete ferric iron chelating proteins, called siderophores[45]. After secretion, a surface-exposed siderophore importer is used to import iron-bound siderophores to harvest iron from the environment. Siderophores tend to be strain-specific so a bacterial species can preferentially and efficiently uptake its own siderophore to harvest iron in a competitive environment[46–49]. Interestingly, *C. jejuni* does not harbor a siderophore biosynthesis pathway, but does encode a promiscuous siderophore importer than can uptake siderophores made by other bacteria[50]. The importer, CfrA, is an important chicken colonization factor, and therefore microbiome derived siderophores contribute to *C. jejuni* success *in vivo*.

### 1.2.3 Genetic Phase Variation in *C. jejuni*

Phase variation is a population-level adaptation mechanism that some bacteria, including *C. jejuni,* rely on to be successful in diverse environments[51–53]. In *C. jejuni*, phase variation is dictated by relatively high-frequency small nucleotide insertion/deletion mutations in homopolynucleotide tracts throughout the genome[25]. These polymorphisms disrupt promoter regions or open reading frames to genetically turn a gene functionally "on" or "off". Considering approximately 30 of these phase variable tracts exist in a *C. jejuni* genome, there are over a billion possible combinations of on/off phase variations. This represents an incredible amount of standing genetic (and

likely phenotypic) heterogeneity within a single *C. jejuni* population. Besides phase variation, standard small nucleotide polymorphisms (SNPs) that naturally occur during bacterial replication are also independently contributing to genetic variation within *C. jejuni* populations. Presumably a subpopulation of cells will intrinsically have the appropriate combination of genes in an on or off phase to succeed after a transition into a new, harsh environment[54]. In this way the standing genetic diversity in a *C. jejuni* population predisposes subpopulations to be successful when the population is exposed to a new environment – for example when transitioning from a piece of chicken into the mammalian intestinal tract[52]. Bacterial cells with genetic variations that increase fitness in the new environment will propagate more successfully and therefore become the major genotype of the transitioned population.

Approximately 30 homopolynucleotide tracts made up of between approximately 5-12 repeated C/G nucleotides are found throughout the *C. jejuni* CG8421 genome. This nucleotide repetition is difficult for the cell to accurately replicate during division and leads to a higher frequency of slip-strand nucleotide insertion/deletion mutations within these tracts. Exhaustive work has identified the frequency of a mutation occurring in phase variable tracts to be once in ~300-3000 replications, depending on the tract[26]. Generally, the farther a tract length is from 7 nucleotides long, the more likely it is to have an altered tract length after division, and the more likely the tract length will become closer to 7 nucleotide long. Interestingly, in eukaryotic cells, repeated nucleotide tracts are thought to theoretically increase in length indefinitely, although in *C. jejuni* they rarely exceed 12 nucleotides and tend to regress towards 7 nucleotides in length on average. The mechanisms dictating these tract length boundaries in *C. jejuni* are unclear, and otherwise, *C. jejuni* has normal rates of genomic mutation and DNA repair fidelity[54].

Notable phase variable genes include genes involved in capsule and LOS structures and flagella modification genes[55]. The role of phase variation in these genes is generally murky because the nuances of surfaces structures are can be difficult to identify. "Locking" a gene into a particular phase to study takes considerable genetic work, and does not control for phase variability of genes that may provide a redundant function.

Although phase variation frequency in a given tract generally requires hundreds of replications to naturally occur under constant environmental conditions[53], exposure to stressors can immediately select for successful variants and eliminate vulnerable variants[52,54]. In this way exposure to a new environment can almost immediately eliminate particular phase variants from a population without the need for growth and replication, and congruently, this would immediately increase the frequency of successful phase variants within that population. This dichotomy can be applied to genomic sequencing of *C. jejuni* populations before and after exposure to a stressor. For instance, the genome of an inoculum population of *C. jejuni* can be sequenced, and the frequency of every phase variation (or any other polymorphism for that matter) can be identified[52]. The inoculum could then be used in animal colonization model, and post-infection isolate populations can be sequenced. Genetic variations that greatly increased in frequency during infection represent genes that are contributing to infection and therefore provide insight into the tools *C. jejuni* is relying on to be successful in the host. Similar experiments have been done in chicken and mouse models of colonization to characterize the extent of phase variation *in vivo*[26,52]. In the work described in Chapter 2, this approach is used to identify *C. jejuni* colonization factors in human infections. Importantly, *C. jejuni* infections can be persistent, and no animal models have been characterized to show similar relapsing, recrudescent infections. Therefore, a human

challenge model is the only available way to study the factors that contribute to *C. jejuni* persistence in mammals.

### 1.2.4 Therapies for *C. jejuni* infection

*C. jejuni* infections are typically self-limiting in immunocompetent hosts[56]. Rehydration is recommended while experiencing symptoms, and if hospitalized, intravenous rehydration can be prescribed. Antibiotics can be prescribed for immunocompromised individuals, or individuals who experience rare systemic disease[57–59]. Severe symptoms, particularly high fever (> 104°F), can also warrant antibiotics[24].

Although antibiotics will quickly eliminate detectable *C. jejuni* in stool, they do not provide relief from the severity/duration of the diarrhea nor abdominal cramping symptoms. Erythromycin is commonly prescribed. Antimotility agents, such as Imodium©, are not recommended as they can make symptoms worse and are associated with a higher risk or mortality[59]. This may be due to a decreased rate of eliminating the bacteria from the intestine via excretion.

Prophylactic antibiotic treatments that typically work for other intestinal pathogens are ineffective for *C. jejuni*[60,61]. Most notably is the failure of Rifaximin to limit Campylobacteriosis in a human challenge model. Rifaximin is a poorly absorbed broad-spectrum antibiotic that is therefore ideal for treating gastroenteritis. Rifaximin prevents shigellosis[60] in a human challenge model and has preventive activity against *E. coli* gastroenteritis, however did not prevent Campylobacteriosis[61].

#### *1.2.4.1 A Controlled Human and Non-human Primate Challenge Models to Test New Therapies*

In a controlled human infection model, a single large scale inoculum is produced and ingested simultaneously by a group of naïve human volunteers[24]. The inoculum is typically a resuspension of bacteria grown on solid-medium, and is a combination of multiple growths and therefore not derived from a single colony. Non-human primates are also susceptible to most human enteric pathogens, and can be challenged in a similar way[62]. Humans are often given bicarbonate solutions to ingest to neutralize bactericidal stomach acid[24], and non-human primates may require combinatorial neutralization treatments (such as addition of ranitidine and/or use of proprietary formulations like CeraVax©) for optimal colonization[62]. As bacteria isolated from human infections are typically used, a lower infectious dose is required to produce symptomatic disease in humans than in non-human primates. For instance, $5x10^5$ bacteria are used to produce Campylobacteriosis in humans, but the new-world primate *Aotus nancymaae* requires $5x10^{11}$ to cause disease. Non-human primates recapitulate human disease better than any non-human primate model, particularly considering the native microbiota do not need altered before challenge[62]. However, the dosage is larger than what would naturally be encountered.

*C. jejuni* strain CG8421 has been characterized for use in controlled human and non-human primate infection challenges[24]. Its utility in these models comes from a lack of GBS-generating LOS structures, so human volunteers are not at risk for the life-threating autoimmune disorder. Azithromycin and ciprofloxacin are used in conjunction to eliminate *C. jejuni* at the end of a human challenge model or if a volunteer experiences severe enough disease to warrant early intervention. These antibiotics were chosen because they effectively kill *C. jejuni in vitro* and *in vivo* at non-toxic levels to the host,

and also accumulate to high concentrations intracellular to be sure invasive bacteria can be reached[63].

### *1.2.4.2 Recrudescent C. jejuni Infections in Humans*

*C. jejuni* has the potential to persist in humans beyond the duration of acute disease. Relapsing - so called "recrudescent" infections – have become a new and significant focus in the *Campylobacter* field[2,9]. *Campylobacter*'s fastidiousness makes isolation from humans more burdensome than for other pathogens, particularly in low resource clinics. This results in an under representation of Campylobacteriosis cases worldwide, and failed detection of recurrent *Campylobacter* infections[2].

However, using multiple diagnostic tests to detect *Campylobacter* in children of low-resource areas, a groundbreaking 2016 study demonstrated the shockingly high rate of *Campylobacter* infection and persistence in children[2]. Out of 2087 children examined across South America, Africa, the Middle East, and South/Southeast Asia, 84.7% of children had a Campylobacter positive stool by 1 year of age. Notably, high campylobacter burdens were associated with periods of growth stunting within the first year, and growth stunting is strongly correlated with lifelong physical and/or cogitative deficits.

Remarkably, 28% of children in low-resource areas were persistently infected by *Campylobacter* during the first year of life, some with asymptomatic colonization and/or potentially continuous re-infection. In adults, the incidence or existence of recrudescent infection was controversial until recently. The success and thoroughness of recent *Campylobacter* human challenge models revealed that almost 20% of adults can present with recrudescent infections[61]. Relapse infections occur about two weeks after the initial acute infection subsides, even if azithromycin and ciprofloxacin antibiotic treatments had

successfully eliminated detectable *C. jejuni* from the stool. The host and pathogen factors that contribute to persistence have not been elucidated, and to date, no small animal models accurate recapitulate human recrudescence. Therefore, human models of infection are the only method to study *C. jejuni* persistence in mammals. The work presented in Chapter 2 identifies bacterial factor that are selected for in human recrudescent infection and elucidates the role of the host in disease severity.

#### *1.2.4.3 Vaccine Development*

Currently there are no vaccines available to prevent *C. jejuni* infections. However, repeat exposures to homologous *C. jejuni* strains offers immunity to future challenges, indicating a Campylobacter vaccine is feasible in principle[29]. However, as described above, our limited understanding of *C. jejuni* pathogenesis and persistence makes rational design of vaccines difficult[64]. Further, standard whole cell or outer membrane vesicle vaccine approaches are concerning given the known risk of generating GBS. However, capsular polysaccharide (CPS) structures do not feature ganglioside mimics, and therefore CPS is an attractive antigen to be used in a vaccine[62,65]. CPS on *C. jejuni* also supports the cell invasion process *in vitro*, and anti-CPS antibodies can hinder this process, indicating a potentially efficacy of anti-CPS vaccines[66]. Generally, though, polysaccharides are poorly immunogenic, and therefore CPS-conjugate vaccines are being pursued[65]. A drawback to these conjugate vaccines are the prevalence of different *C. jejuni* capsule serotypes, with approximately 60 different serotypes identified. However, a multivalent vaccine currently in clinical trials could protect against over 30% of known serotypes, and at least demonstrate the efficacy of such conjugate vaccine approaches in *C. jejuni*.

Therapy development would benefit from knowledge of additional targets that contribute to human infection. Currently, chicken colonization models (a native host) are used to identify colonization factors, but the zoonotic lifestyle of Campylobacter makes extrapolating these data to human infection perilous. The work described in Chapter 2 identifies genetic factors that are selected for during human disease, along with genes that are highly expressed during human infection. Rational therapy design may benefit from considering these factors when choosing antigens for future vaccines.

## 1.3 Enterotoxigenic *E. coli* (ETEC): A Global Cause of Traveler's Diarrhea

### 1.3.1 Impact and Sources of ETEC Infections

Enterotoxigenic *Escherichia coli* (ETEC) is a gram-negative bacterial intestinal pathogen that infects 200 million people every year[1,3,67]. It was originally described in the early 1970s and infection typically results in acute watery diarrhea which can potentially result in lethal dehydration[4,68–70]. Assessments to define diarrhea vary amongst clinicians, however it is common to consider three loose stools within 24 hours to be diarrhea. Additional symptoms of ETEC infection can include abdominal cramping and vomiting[4]. However, vomiting is rare compared to the frequency of diarrhea-associated symptoms. ETEC diarrhea is typically an acute, profuse, watery diarrhea and therefore is often described as "cholera-like"[4,71]. Unsurprisingly, then, is the presence of a cholera-like toxin in some strains of ETEC[69]. These virulence factors are discussed in Chapter 1.3.2.

Classic ETEC symptoms of acute diarrhea often last between 4 and 14 days. These symptoms are often debilitating, leading to low productivity during illness[72]. A 2010 study determined that ETEC infections account for an annual loss of 6.8 million disability-adjusted life years of human productivity[72]. Further, pervasive ETEC diarrhea occurs in some patients and can last 30 days or more and require treatment with

antibiotics[73,74]. Given an approximation that the average duration of ETEC illness is 7 days, and considering 200 million people every year are infected by ETEC, one can estimate that over 1 billion days of human productivity are lost annually to ETEC infections. This is a loss of approximately 2.7 million years' worth of societal output – inarguably a devastating loss - and a strong motivator to better understand and treat ETEC infections.

Although infection of adults in high-resource areas is rarely deadly, young children and infants in low resources areas are at a high risk of death due to their naïve immune system and propensity for rapid dehydration[75]. The duration of diarrhea can also result in malnutrition in infants which is associated with stunted growth periods. Notably, similar stunted growth phases during the first year of life in low-resource areas is associated with life long physical and/or cognitive deficits[2,72].

Low-resource areas throughout the world, including most countries in Latin America, Africa, south/southeast Asia, have the largest burden of ETEC infection[1,72]. ETEC infections are considerably rarer in high-resource areas like the United States and Europe which utilize effective and widespread public water treatment. Therefore, food-borne outbreaks are the most typical source of rare ETEC infections in these developed regions[74,76].

Overall, children in low-resource areas account for most ETEC infections, and those that survive gain natural immunity to their local ETEC strain(s)[75]. International travelers to endemic countries are also at extreme risk of ETEC infection with an estimated 10 million becoming ill each year. Accordingly, as many as two thirds of visitors to tourist attractions in endemic areas have been reported to contract ETEC infections[77]. Although boiling water and eating cooked food can decrease the risk of

contracting ETEC, it can be difficult for travelers to prepare their own food or to properly ensure their purchased food and drink was appropriately prepared.

The United States military is also burdened by ETEC infection, as an estimated 25-50 percent of troops deployed in developing countries will have ETEC-associated diarrhea during combat or training[78,79]. As reported by Bloomberg magazine in December 2015, the military estimates they lose over 1 million days of soldier activity due to diarrheal infections each year. This has spurred significant research endeavors within the United States Navy Medical Research Center to better understand ETEC infections[80]. Their efforts to develop ETEC vaccines, along with the treatment endeavors of other groups, in detailed in Chapter 1.3.4.

ETEC infections are difficult to prevent in endemic areas because contaminated water is a major source of ETEC. When water purification resources are inadequate it can be difficult to contain ETEC contamination, and congruently, surface waters in afflicted areas are frequently found to contain ETEC[81–83]. This is exacerbated by the vast amount of infected diarrhea produced by afflicted humans which can easily contaminate environmental sources of water. When this contaminated water is used recreationally, consumed as drinking water, or used to wash produce and dishes, infection is almost certain in naïve consumers[81].

ETEC endemics are often seasonal, likely due to people associating with contaminated recreational waters more frequently in warmer seasons[76,84]. However, ETEC can also be found indoors. The presence of ETEC biofilms within the indoor water reservoirs of homes in low-resources areas are associated with seasonal epidemics[84].

ETEC infections are frequently referred to as “travelers’ diarrhea,” and certainly it is a leading cause of such diarrhea. However, it is also a deadly infection in children, and infants that do survive may have lifelong physical and/or cognitive deficits as a result[1,2].

Although survivors of ETEC infection gain a natural immunity to their local ETEC strain, this often does not provide protection to ETEC strains found in other areas. Therefore, traveling to another region with a different endemic strain can lead to infection, and immunogenicity to ETEC is discussed in Chapter 1.3.4. The burden on travelers extends to military personnel who are deployed around the world, and infections in soldiers has spurred the Navy in investigate new ETEC vaccines used in human clinical trials[78–80]. Samples collected from these human challenges are used in this dissertation work and are described in Chapter 3.

### 1.3.2 ETEC Virulence Factors

ETEC can use a trio of plasmid-borne "classical" virulence factors to cause disease in mammals[70,85,86]. Bacterial adhesion to host cells is a required colonization factor for almost all bacterial pathogens. Adhesins are surface exposed pilus-like structures that allow bacteria to physically adhere to surfaces or host cells. At least one adhesin is used by ETEC to adhere to the intestinal epithelium to cause symptomatic disease[87]. This allows for 1) the bacteria to attach and associate with the host intestinal epithelium 2) ensure affective delivery of secreted toxins directly to host epithelial cells 3) resist the flushing action of the diarrhea that is induced by these toxins. ETEC adhesion are discussed in Chapter 1.3.2.1.

Beyond adhesins, ETEC is defined by the enterotoxin genes it harbors. Enterotoxins are proteinaceous secreted toxins that induce an efflux of water from the cytoplasm of host epithelial cells into the lumen of the intestine[68,69]. This efflux of water is the source of the characteristic watery diarrhea. The prototypical ETEC strain H10407, which was originally isolated from an adult with cholera-like symptoms in 1973[85], harbors both a heat-labile enterotoxin and two heat-stable enterotoxin homologs

and is examined in this dissertation. Enterotoxigenic heat-stable (ST) and heat-labile (LT) toxins, although structurally and enzymatically different[88,89], both induce diarrhea. Despite many ETEC isolates harboring multiple enterotoxins[90,91], only one is required to induce symptomatic disease when accompanied by a functional adhesin[86]. These enterotoxins are described in more detail in Chapter 1.3.2.2.

These classical ETEC virulence factors (adhesins and enterotoxins) are found on so-called virulence plasmids within ETEC strains[85,92]. The number of plasmids and virulence factors varies per ETEC isolate, however the prototypical ETEC strain H10407 harbors 4 virulence plasmids (p948, p666, p52, p58) and four classical virulence factor homologs (an adhesin called CFA/I, two ST homologs, and LT)[93].

### *1.3.2.1 ETEC Adhesins*

Approximately 25 different adhesins have been identified in ETEC strains isolated from humans around the world. However, up to 50% of clinical isolates test negative for any known adhesin type despite having adhesive ability *in vitro* and causing human disease[94]. This indicates the requirement for adhesion can be fulfilled by a variety of apparatuses and that significantly more investigation is warranted into adhesin diversity in ETEC. This dissertation will focus on the prototypical adhesin, Colonization Factor Antigen I (CFA).

The CFA adhesin is found in 30-40% of ETEC isolates, including in strain H10407, and is well characterized[75]. In H10407, the operon *cfaABCE* encodes the main structural subunit of the CFA adhesive pilus (CfaB), the adhesive tip found at the end of the pilus (CfaE), and the outer membrane usher and periplasmic chaperone proteins (CfaC and CfaA, respectively) which build the extracellular pilus[87,95]. The structures of CfaB and CfaE has been determined – both of which are barrel shaped proteins with 7-

stranded β-sandwich folds of the Ig-family[87]. Surprisingly, there is minimal contact between CfaE and CfaB, with only 4 hydrogen bonds enabling CfaE association with the last CfaB subunit of the pilus.

A eukaryotic membrane receptor for CfaE-mediated adhesion has not been identified. However, ETEC that infect pigs harbor porcine-specific adhesins[96] which have known glycoprotein receptors[97]. Surface exposed mucin-type glycoprotein (IMTGP) on porcine brush border intestinal epithelial cells is bound by the porcine ETEC adhesin K88. Pigs that do not produce this receptor ($IMTGP^-$) are less susceptible to symptomatic disease when challenged with a $K88^+$ ETEC strain than $IMTGP^+$ pigs[98]. Interestingly, porcine ETEC strains that rely on the K88 adhesin do not appear to cause disease in humans, although porcine homologs of ST and LT are found in human ETEC isolates that cause disease[99].

The *cfa* operon is found on the ETEC H10407 virulence plasmid p999, along with the regulator of the operon, *cfaD*[100,101]. *cfaD* is generally considered constitutively expressed and is a DNA binding protein which activates expression of the *cfaABCE* operon by binding to its promoter region[102]. Although CfaD is unique to ETEC, it is part of the AraC-family of virulence transcriptional regulators which are found in many pathogens and bind asymmetrically to AT-rich sequences in promoter regions of virulence genes[103,104]. This binding can displace the repressive binding of the histone-like nucleoid-structuring protein (H-NS) at these promoter regions[102].

A rich media that was used to grow ETEC soon after its discovery ensures robust expression of the CFA adhesin in strain H10407, and therefore has the namesake "CFA media" and is now a standard growth media when working with ETEC *in vitro*[80]. Expression of *cfaD*, and therefore expression of the *cfa* operon, was recently shown to be increased by exogenously added bicarbonate ($HCO_3$)[100]. Bicarbonate has been shown to

influence virulence factor expression in a number of pathogens, likely due to providing a more accurate approximation of *in vivo* $CO_2$ concentrations during laboratory culture growth[105,106]. Beyond this work, however, little is known about the regulation of *cfaD* expression.

Beyond activating expression of the *cfa* operon, CfaD activates expression of itself, along with a few "non-classical" virulence-associated genes which are discussed in Chapter 3[107]. It is worth explicitly noting, however, that cfaD does not regulate the expression of either ST or LT[100], which are discussed in the following section.

#### *1.3.2.2 Secreted Toxins*

H10407 harbors both heat-stable (ST) and heat-labile (LT) enterotoxins[93]. The genes *sta1* and *sta2* are two ST enterotoxin homologs found on virulence plasmids p666 and p948 respectively[92]. Sta1 is a secreted 18 amino acid peptide that is also found in ETEC strains associated with porcine infection, and therefore is sometimes called STp. Sta2 is a secreted 19 amino acid peptide is that exclusively found in ETEC isolated from human infections and therefore is sometimes called STh. As both of these toxins are discussed in the context of the human isolate H10407 they are referred to as Sta1 and Sta2 in this dissertation. It is not uncommon for ETEC isolates to harbor multiple enterotoxins and at least three other STh variants have been described[108].

The Sta genes encode for an initial peptide of 72 amino acids in length which is eventually processed to produce a short (about 18 amino acid long) peptide for secretion[109]. The peptides contain a 19 amino acid pre-region, which is a signal peptide for secretion into the periplasm. This pre-region is cleaved off via signal peptidase. Once in the periplasm, disulfide bonds important for toxin function are formed. Then ~44 amino acids are cleaved off the n-terminal of the peptide, leaving 19 (Sta2) or 18

(Sta1) amino acids that represent the mature, functional, secreted toxins[109]. Secretion then occurs in a TolC-dependent manner.

In the host, secreted ST toxins bind to the transmembrane guanylyl cyclase C receptor (CG-C)[110,111] on the brush boarder membrane of epithelial cells. CG-C plays an important role in controlling the fluid content of the intestine by regulating electrolyte and associated water efflux into the intestine[111]. Binding of ST to the extracellular domain of CG-C activates the CG-C intracellular domain to catalyze the production of cyclic guanosine monophosphate (cGMP)[112]. Elevated levels of intracellular cGMP activate cGMP-dependent protein kinase PKGII, which in turn activates a transmembrane chloride channel (cystic fibrosis transmembrane conductance regulator, CFTR) via phosphorylation[112]. Activated CFTR exports chloride ions ($Cl^-$) into the lumen of the intestine which at increased concentrations induces an osmotically induced efflux of water from host epithelial cells into the lumen[113].

ST-induced cGMP accumulation in host cells can stimulate electrolyte efflux via two additional pathways. The first is by inactivation of phosphodiesterase 3 (PDE3) which results in accumulation of intracellular cyclic adenosine monophosphate (cAMP)[114]. Increased intracellular cAMP concentrations activates protein kinase A (PKA) which also stimulates CFTR via phosphorylation to export $Cl^-$ into the lumen[115]. Finally, PKA also inhibits a $\underline{N}a^+/\underline{H}^+$ exchanger membrane channel (NHE3) which imports $Na^+$ from the lumen via exchange of $H^+$ export into the lumen[116]. By inhibiting Na+ import into the host cells, the electrolyte concentration in the lumen is further elevated, leading to increased osmotic stress that results in water efflux from the cytoplasm into the lumen. Notably, CFTR null mice are viable, resistant to ST-induced water efflux in the intestine, and resistant to $ST^+ LT^-$ ETEC colonization[117].

LT is a hexameric secreted enterotoxin found in both porcine and human ETEC isolates[89,99]. It is an exemplary $AB_5$ toxin, where 5 "B" protein subunits (LTB) and 1 "A" protein subunit (LTA) associate make up the functional holotoxin. The genes for LTA and LTB are named *eltA* and *eltB* in H10407 and are overlapping in an operon (*eltAB*) on virulence plasmid p666[93]. It is homologous in structure, function, and nucleotide/amino acid sequence to cholera toxin from the intestinal pathogen *Vibrio cholerae*[118].

Evolutionary sequence analysis estimated E. coli acquired LT via a horizontal gene transfer event from *V. cholerae* approximately 130 million years ago. The eltAB operon is approximately 80% homologous to the cholera toxin operon, *ctxAB*. This most divergent domains between the two toxins are the N-terminal signal peptide sequences which target the proteins to the periplasm. This sequence is cleaved off before secretion by signal peptidase and therefore is consistent with a domain that can tolerate sequence variation. After reaching the periplasm, LTA and LTB spontaneously oligomerize to form LT with appropriate stoichiometry[119] and is secreted via the *gspC-M* type II section pathway[120]. Secreted LT can travel freely in the extracellular milieu, or is also capable of associating with bacterial outer membrane (OM) or extracellular outer membrane vesicles (OMVs) via interaction with membrane anchored bacterial sugars[121].

Once secreted, LT functions identically to cholera toxin to induce diarrhea. First, LT binds to a monosialoganglioside GM1 receptor on the surface of epithelial cells and is subsequently endocytosed[122]. The LTA subunit then enables ADP-ribosylation of the Gsα subunit of the host G-protein in the endoplasmic reticulum[123]. This activates G-protein to activate adenylyl cyclase which cleaves ATP to cAMP. Increased levels of intracellular cAMP activates protein kinase A, which results in similar electrolyte imbalance via NHE and CFTR as discussed with ST. The differences in the host

response to free LT vs OMV or OM associated LT is currently an active area of investigation in ETEC pathogenesis[124].

### 1.3.4. Available Treatments and Therapy Development for ETEC Infections

#### *Current Treatments*

Generally, ETEC infections are self-limiting and don't require antibiotic intervention[75,125,126]. Hydration with clear liquids is recommended until diarrhea subsides. Prophylactic treatment with the antibiotic Rifaximin can be prescribed to at-risk travelers who may not be able to tolerate a bout of infectious diarrhea due to other health conditions[127]. Although Rifaximin is an effective prophylactic treatment for *E. coli* and *Salmonella* infections it is not affective against *C. jejuni* infections[61]. Bismuth subsalicylate (Pepto-Bismol©) and anti-intestinal-motility agents (Imodium©) are effective at decreasing the number of loose stools and lessening the severity of diarrhea-associated cramps[128]. If antibiotic intervention is required in severe cases, azithromycin is typically prescribed, as it is also affective against other intestinal pathogens (including *C. jejuni*) and there is low resistance amongst ETEC isolates[80].

#### *Vaccine Development*

Vaccines are a means to induce protective immunity against a pathogen to prevent future infections. Travelers to afflicted regions are at high risk for ETEC infection, and therefore a vaccine that can provide protection from ETEC infection before traveling is an attractive area for therapy development. ETEC's classical virulence factors (adhesins, LT, and ST) are all highly immunogenic, and host antibodies against these factors provide protection in animal models of infection. This explains why after becoming infected with a local ETEC strain, reinfection does not occur. However, ETEC isolates

from around the world have shown there is significant diversity amongst virulence factors, particularly the adhesins, and this explains why standing immunity to a local ETEC strain may not provide protection against a strain from another region of the world[94]. There is lesser diversity seen amongst ST and LT homologs so they are used more frequently as antigens during vaccine development[108].

To overcome the geographic diversity seen in ETEC, the most promising vaccines are multivalent. For example, a vaccine that can induce immunity against the most common adhesins seen in ETEC (CFA/I and CS1-6) and LT could provide protection against 80% of all known ETEC strains[80].

Controlled human infection models (CHIM) are affective ways of testing the efficacy of ETEC vaccines. In a CHIM, human volunteers simultaneously ingest a portion of a lab grown ETEC culture and become infected in a clinical setting. Using a double blinded approach, volunteers can be administered either a test or sham ETEC vaccine prior to challenge, and the efficacy of the test vaccine can be determined by comparing the severity of diarrhea experienced by volunteers who were given the test vaccine vs the sham control vaccine[80].

Recently a human clinical trial and CHIM was preformed to test the efficacy of vaccine using CfaE and inactivated LT as the protective antigens (ClinicalTrials.gov Identifier: NCT01922856). In the trial, human volunteers were challenged with ETEC H10407 after being vaccinated against CfaE/LT or with a control sham vaccine. The results of the trial have yet to be made public, however this dissertation uses samples obtained from these volunteers to investigate ETEC directly in human infection samples.

The CHIM provides a number of advantages to studying ETEC directly in *in vivo infection* samples than collected diarrhea samples from naturally acquired ETEC infections. First, all patients consume a portion of the same inoculum simultaneously.

This enables direct comparison of bacteria across infected samples without confounding differences between strains or time post-inoculation. Second, volunteers are pre-screened for the naivety to ETEC infections, so confounding pre-existing immunity to the infected strain need not be considered. Finally, the H10407 strain used in the recent CHIM is well characterized[71,80,85,126], including the full genome sequence[93] and extensive research on virulence factor production and regulation[129]. This allows conclusions drawn from this study to be immediately considered within the context of decades of previous work.

## 1.4 Characterizing Bacterial Adaptations Reveals Important Survival Mechanisms

Organisms must adapt and respond to changes in their environment in order to survive. Bacteria are no exception, and bacterial pathogens must be fit to survive in a wide range of environmental conditions. Two major ways bacteria adapt to their environment are 1) sensing of important environmental cues and responding to them via changes in gene transcription[130] or by 2) a genetic response - direct selection for particular bacteria that are more fit to survive - which is due to genetically encoded differences between bacterial cells[131]. Each of these adaptation methods, and how we can characterize them to reveal the tools bacteria used to survive or cause disease, are discussed below.

### 1.4.1 Using Next-generation RNA Sequencing to Characterize Transcriptional Adaptive Responses in Bacteria.

*C. jejuni* is able to survive in diverse environments - as it survives in mammalian intestines, avian intestines, milk, environmental water, in refrigerated conditions and/or on raw poultry, and within host cells[2,12,132]. It must sense changes between these environments and adapt accordingly in order to survive in these diverse conditions. For example, *C. jejuni* senses bile as a que to promote survival in the intestine[133]. Bile is an

important emulsifier in the intestine and *C. jejuni* harbors a transcriptional regulator named CmeR which is directly bound by bile acids[134]. In the absence of bile, CmeR binds to the promoter regions of over 20 genes and represses their expression[130]. The CmeR regulon includes a bile acid efflux pump, CmeABC, which promotes resistance to bile acids[133]. When bile is present in the environment, such as within a host intestine, CmeR becomes bound by bile acids which promoters a conformational change in the protein that releases the protein from DNA[134]. This relieves CmeR repression and activates expression of the CmeR regulon, including the CmeABC bile acid efflux pump. In this way *C. jejuni* senses an environmental que to transcriptionally respond to promote survival within a host. Studies looking at pathogen transcriptional changes during mammalian infection have revealed similarly important survival mechanisms – notably in human uropathogenic *E. coli*[135]– and have helped established examining transcriptional changes as a way to identify important adaptive mechanisms of survival during infection.

The totality of gene transcription in a cell is called the transcriptome, which denotes the expression level of every gene in the genome[136]. Next-generation, high-throughput nucleotide sequencing of total RNA from bacteria can be used to examine global transcriptional changes between bacteria grown in different environmental conditions or between different bacterial strains/isolates.

To accomplish this, RNA is extracted from bacteria using guanidinium thiocyanate-phenol-chloroform extraction, or similar phenol-based extraction methods, which use chemical phase separation to separate RNA from genomic DNA and protein[137]. Addition of DNAse enzyme to extracted RNA can be used to eliminate contaminating genomic DNA form the extraction. As over 90% of RNA is unfluctuating ribosomal RNA (rRNA)[138], removing rRNA makes it easier to characterize changes in non-ribosomal RNA transcript levels[139]. So, DNA-free RNA can be incubated with ribosomal

RNA quenching probes, which bind to and physically sequester rRNA, so that only non-rRNA remains in additional steps.

After DNA and rRNA have been removed, a library of complementary DNA (cDNA) is made using a reverse transcriptase enzyme and polymerase chain reaction amplification. This library can then be prepared for Next-generation sequencing using proprietary platforms, like Illumina[140]. This sequencing produces millions of reads of the cDNA which represents the original transcriptome of the bacterial sample population. These reads can then be bioinformatically mapped to the known, previously sequenced bacterial genome. For a standard transcriptome analysis, each read is mapped to its homologous location in the genome, and the number of reads that map to annotated regions of the genome are counted, per gene and/or annotation. To compare if a particular gene had more mapped reads during growth one environmental condition (or perhaps expressed differently between bacterial isolates) and therefore showed differential expression between samples, the read mapping must be normalized[141]. A common method for this is to normalize the number of mapped reads to gene by the length of the gene and the total number of reads that were sequenced for that sample. A popular method for this normalization is called RPKM, for Reads mapped Per Kilobase of a gene per Million reads sequenced[141]. This helps remove bias in gene length (longer genes will have more reads map to them without necessarily having higher transcription levels) and normalizing for samples that have more or less total sequenced reads (more sequenced reads will result in more mapped reads despite transcription levels). Finally, RPKM values for every gene in the genome can then be compared between samples to determine which genes have expression changes between samples. A statistical test, such as Baggerly's test[142], should be applied to these data with a false-discovery rate (FDR) corrected p-value to identify statistically significant changes in gene expression.

Typically genes with a least a 3 fold change in expression with a FDR p-value of < 0.05 are considered significantly changed between samples, although this arbitrary and may not be appropriate for all analyses. Notably, the biological relevance of a particular fold change is likely different per gene, and therefore significance cutoffs should be considered in the context of the scope of the work.

Work in this dissertation used a similar method of RNA-sequencing to determine the *C. jejuni* and ETEC transcriptomes directly in human infection samples compared laboratory growths. Additional complications arise when working with complex samples such as human feces. First, additional rRNA removal steps must be included to remove eukaryotic rRNA. Second, partially digested food in feces is a potent inhibitor of DNAse during DNA removal. To ameliorate this, a larger volume (~50ml vs 500ul) was used during guanidinium thiocyanate-phenol-chloroform extraction which produced greater physical separation of DNA from RNA during extraction to limit DNA contamination. Finally, only a small proportion of RNA extracted from infected feces is from the target pathogens, so billions of reads needed to be sequenced to get appropriate coverage of the genome. Strict mapping parameters - such as using 100% exact homology mapping to the reference genome - were used to eliminate non-specific transcripts from other bacteria influencing the results. These methods are described in more detail in the final chapter of this dissertation.

### 1.4.2 Using Next-generation Genomic Sequencing to Reveal Mechanisms of Pathogen Success During Infection

A variety of mechanisms can induce genomic mutations in bacteria, including errors made by DNA polymerases during replication and DNA damage and repair mechanisms. Some genomic mutations are silent i.e. they do not alter the amino acid

sequence or expression of any gene or affect a regulatory element. These silent mutations have no observable influence of the fitness of the bacteria. However, influential mutations, which do change protein coding sequences or regulatory elements can alter the fitness of the bacteria that harbor them[131]. Influential mutations can either increase or decrease the fitness of bacteria[52]. Bacteria with mutations that improve fitness will propagate more successfully than their neighbors and therefore eventually increase infrequency within the bacterial population. Bacteria that harbor mutations that decrease fitness will allow neighboring bacteria to outcompete them, and therefore detrimental mutations will decrease in frequency in a population. Some mutations are immediately lethal, for instance if they inactivate an essential gene required for replication[143], and therefore they are immediately lethal to the bacteria and are therefore rarely detected within the populations.

This concept of genetic variation leading to competing fitness over time is the basis of evolution and can be used to identify genetic elements that enable bacteria to be successful. In the past, various methods of identifying genomic loci responsible for bacterial phenotypes were used, including phage transduction and conjugation. With the advent of high-throughput DNA/genomic sequencing entire bacterial genomes can be sequenced to identify all genetic differences between bacterial isolates and has been a boon for understanding bacterial diversity both between and within bacterial populations[144].

Importantly, genomic sequencing can be used to identify and track the spread of antibiotic resistance loci, precisely survey pathogen transmission between patients, and understanding the genetic selection pressures in complex environments[6,7,145–147]. The host environment is remarkably complex - intestinal pathogens must survive in the context of a dense commensal microbiota, severe nutritional limitations, host digestive compounds,

ingested food, and an active host immune response. Therefore, investigating genetic adaptations that are selected for *in vivo* can give unparalleled insight into the tools bacteria use to survive *in vivo*, as these adaptations could only arise while within this complex environment and would difficult to anticipate or replicate in the laboratory.

This dissertation presents genomic sequencing of bacterial isolate populations before and after human infection. Bacteria that naturally harbored beneficial genetic mutations outcompete their neighbors during infection and therefore increase in frequency in the infection population when compared to their frequency in the inoculum[26]. Identifying these changes in frequency reveals the genetic adaptations that were selected during infection and therefore identifies the genes or "tools" the bacteria relied on to be successful in the human host.

# Chapter 2: *Campylobacter jejuni* transcriptional and genetic adaptation during human infection

## 2.1 INTRODUCTION

Studying bacterial evolution during infection exposes how pathogens adapt and survive in the host. High-resolution whole genome sequencing enables accurate tracing of pathogen transmission between patients, identification of antibiotic resistance loci, and understanding of selective pressures *in vivo*[6,145–149]. These developments have led to the identification of pathogen genetic variants that predict the success of treatments and guide therapy design. Such studies reveal the extent of bacterial adaptability *in vivo* while regarding the genetic fitness compromises that would only arise in the complex host environment.

Some bacteria adapt to new hosts and environments using a common population-level adaptation method known as phase variability[26]. Phase variation is driven by relatively high frequency genome mutations in homopolynucleotide tracts which result in a functionally "on" or "off" phase of affected gene(s)[26]. Subpopulations of on/off gene variants are intrinsically more fit for a new environment and therefore increase in frequency in a population as they outcompete less fit variants. This adaptive process genetically regulates virulence factors and is important for pathogen colonization and immune evasion[51,52,150].

---

[1] Portions of this chapter have been previously published: Crofts, A., Poly, F., Ewing, C., Kuroiwa, J., Rimmer, J., Harro, C., Sack, D., Talaat, K., Porter, C., Gutierrez, R., DeNearing, B., Brubaker, J., Laird, R., Maue, A., Jaep, K., Alcala, A., Tribble, D., Riddle, M., Ramakrishnan, A., McCoy, A., Davies, B., Guerry, P. & Trent, S. Campylobacter jejuni transcriptional and genetic adaptation during human infection. Nature Microbiology 3, 494–502 (2018). Crofts conceived, designed, preformed and interpreted genetic and transcriptomic analyses and wrote the paper. Guerry, Trent, and Davies oversaw this study and edited the manuscript. All other coauthors contributed to primate studies.

To fully understand phase variable populations, whole genome sequencing can be used to identify the frequency of every "on" or "off" gene variant at the population level and then identify influential variants that are selected for during adaption to new host environments[151,152].

*Campylobacter jejuni* is a phase variable Gram-negative intestinal pathogen that causes bloody diarrhea, fever, and abdominal pain in humans. Notable secondary sequelae include Guillain-Barré syndrome, a potentially fatal paralytic autoimmune disorder. *C. jejuni* is a prevalent commensal bacterium in the intestinal tracts of chickens and other livestock, and consumption of contaminated animal products makes *C. jejuni* a leading cause of food-borne bacterial diarrhea worldwide. In low-resource areas, asymptomatic and sometimes persistent *Campylobacter* infections are common in children younger than 1 year old and correlate with stunted growth and therefore life-long physical and cognitive deficits[2]. In line with our understanding of *in vivo* pathogen evolution, small animal models of *C. jejuni* infection show host-passaged isolates are more fit in the host than the initial inoculum, and this genetic advantage remains even after multiple passages *in vitro*[153]. Here we use samples from a controlled human infection model to characterize the *C. jejuni* response to the human host. High-resolution genomic sequencing was used to track pathogen adaptation and phase variation *in vivo*, from inoculum through acute and persistent disease. In the challenge, a single liquid inoculum of *C. jejuni* strain CG8421 was used to simultaneously infect a group of human volunteers, making this study a highly-controlled pathogen adaptation experiment in humans. The challenge was designed to evaluate the prophylactic efficacy of rifaximin on *C. jejuni* infection in humans and is described elsewhere (ClinicalTrials.gov Identifier NCT02280044)[61]. Briefly, rifaximin treatment had no effect on the primary clinical outcome of campylobacteriosis. To be released from the trial, volunteers were treated

with azithromycin and ciprofloxacin and produced feces negative for *C. jejuni*. However, five volunteers experienced recrudescent[154] infections. The clinical trial outcomes enabled us to compare pathogen gene variants that were selected for during both acute and recrudescent[154] infections, as well as in volunteers who experienced severe disease or received prophylactic antibiotic treatment. We found variants of genes involved in host cell invasion, bile sensing, and flagella modification are selected for in recurrent human infections. To validate our approach, we performed a similar analysis in a symptomatic non-human primate infection model. Beyond defining genetic adaptation in primate hosts, we also used RNA-sequencing to determine the *C. jejuni* transcriptome directly in human infection samples.

## 2.2 RESULTS

### 2.2.1 Campylobacter Gene Expression in Human Infection Samples

Responding to the host environment via transcriptional changes is essential for *C. jejuni* colonization and infection[155,156]. To identify the transcriptional adaptations that occur in a human host we determined the *C. jejuni* transcriptome directly in infected human feces of three volunteers. Compared to *in vitro* microaerobic growth on blood agar, 264 genes were differentially regulated *in vivo* by at least 3 fold (false-discovery rate adjusted p-value <0.05) (Fig. 2.1a, Sup. Table 1). A similar number of genes were differentially regulated compared to growth on Mueller-Hinton agar and in Mueller-Hinton broth (Fig. 1a). No matter which comparison is made, a conserved group of 110 genes were differentially regulated *in vivo*. These gene products include diverse characterized colonization factors, including Peb1A, PldA, CsrA, and the capsule polysaccharide transporter KpsM[132,157–159]. These data also reveal only between 10 and 15

percent of annotated open reading frames show significant differential regulation in infected human feces compared to growth in standard laboratory conditions.

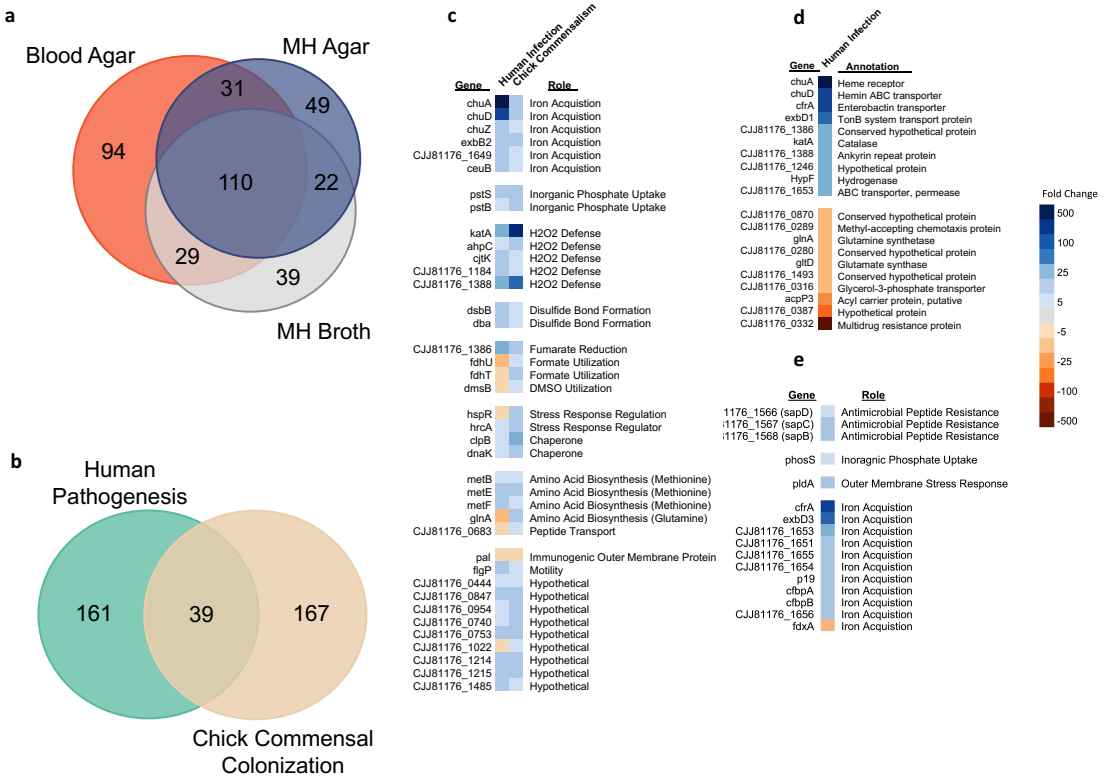



Figure 2.1: The *C. jejuni* Transcriptome in Human Infection and Chick Commensalism.

**a**, Differentially regulated genes *in vivo*. A Venn diagram showing the number of differentially regulated *C. jejuni* CG8421 genes (fold change >|3|, FDR p-value <0.05) in the infected feces of three volunteers compared to laboratory control growths which each represent three biological replicates. **b**, Human infection versus chicken commensalism transcriptomes. A Venn diagram showing the number of differentially regulated (fold change >|3|, FDR p-value <0.05) gene homologs *in vivo* between *C. jejuni* CG8421 infected human feces (3 samples) and a previously published RNA-seq transcriptome of *C. jejuni* 81-176 in the chicken cecum, which represents 3 pools of 5 infected chick ceca each. Both data sets use mid-log phase growth in Mueller-Hinton broth biological triplicates as the *in vitro* lab comparison. **c**, Conserved differential gene regulation across hosts, lifestyles, and strains. The 39 differentially regulated homologs from Fig. 1b conserved between human infection and chicken commensalism with transcriptional fold changes noted. **d**, The top most upregulated and downregulated CG8421 genes in human infection samples. **e**, Notable genes uniquely differentially regulated in human infection samples.

Although primates can experience disease symptoms during infection, *C. jejuni* has a commensal lifestyle in the natural bird host. To relate transcriptional adaptations that occur during infection and commensalism, we compared our data to a previously published RNA-sequencing data set of *C. jejuni* strain 81-176 in the chick cecum (Fig 1.1b, Sup. Table 2)[160]. We identified a core set of 39 conserved genes that are differentially regulated *in vivo* across hosts, colonization lifestyles, and strains. These gene products share strikingly similar roles, including iron acquisition, inorganic phosphate uptake, and protection from peroxide stress (Fig. 2.1c). Epithelial cell production of hydrogen peroxide plays an important role in mucosal immunity against *C. jejuni*. Hydrogen peroxide-mediated inactivation of the *C. jejuni* outer membrane tyrosine kinase Cjtk leads to a decrease in capsule production and therefore decreased virulence[161]. Here we see the importance and conserved nature of this response as *cjtK* along with two major hydrogen peroxide scavengers, *katA* and *ahpC*, are upregulated during both chicken commensalism and human infection. Finally, 7 of these 39 genes showed discrepant expression patterns. Notably, the *fdhTU operon*, which is important for formate dehydrogenase activity[162], is oppositely expressed between the human and chick data.

The most upregulated genes in human infection samples represent two distinct iron acquisition pathways. The *chuA* and *chuD* hemin uptake genes, and *cfrA*, a ferric-enterobactin uptake gene, are all upregulated >100 fold *in vivo* (Fig. 2.1d). *chuA* is not required for chicken colonization[163] and is modestly upregulated in the chick cecum (<10 fold increase *in vivo*), although it is highly upregulated in human infection samples (>1000 fold increase *in vivo*). There are 161 genes that are differentially regulated specifically in our human samples and not in the chick cecum (Fig. 2.1b) and includes the highly upregulated *cfrA*. *C. jejuni* does not produce siderophores, but CfrA enables

import of a wide array of siderophores made by other organisms and, unlike *chuA*, is required for chicken colonization[50]. This suggests microbiome derived siderophores and *cfrA* play an important role in *C. jejuni* infection of humans. Interestingly, components of the Sap antimicrobial peptide resistance efflux pump, which is required for chicken colonization[164], are also uniquely upregulated in human infection (Fig. 2.1e).

### 2.2.2 Genome Variant Selection During Human Infection

To identify genetic variants that are selected for during human infection we examined the genome of *C. jejuni* before and during acute and persistent human infections. We sequenced the genomes of 49 *C. jejuni* CG8421 infection isolate populations from 14 volunteers during acute and recrudescent infections and compared them to the inoculum genome (Fig. 2.2a, Fig. 3).

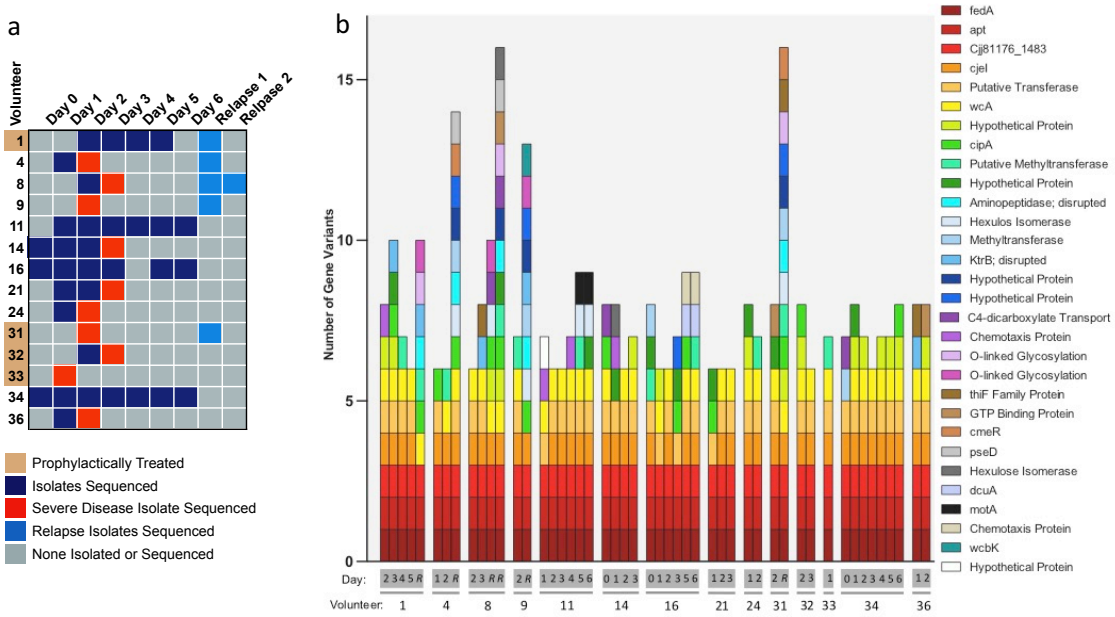



Figure 2.2 *C. jejuni* Genomic Variants are Consistently Selected for During Human Infections

**a**, Infection populations sequenced. Whole genome sequencing was performed on 49 infection isolate populations, taken from 14 volunteers across 6 days of primary infection. 5 volunteers relapsed after the clinical trial, and those infection isolate populations were also sequenced. Some volunteers experienced severe disease on the noted days (red) and therefore received early antibiotic intervention. Some volunteers received exploratory prophylactic treatment with the antibiotic Rifaximin as noted. Sequencing details are noted in Fig. 2.3. **b**, Genomic variants detected per volunteer isolate population, noted by the genes the variants affect. The number of genomic variants detected per infection population is noted on the y-axis, and the x-axis denotes the day and volunteer the sample was taken from, with R denoting a relapse sample. The corresponding genes are listed by their gene name or annotation when considering homologs across CG8421, 11168, and 81176 *C. jejuni* strains.

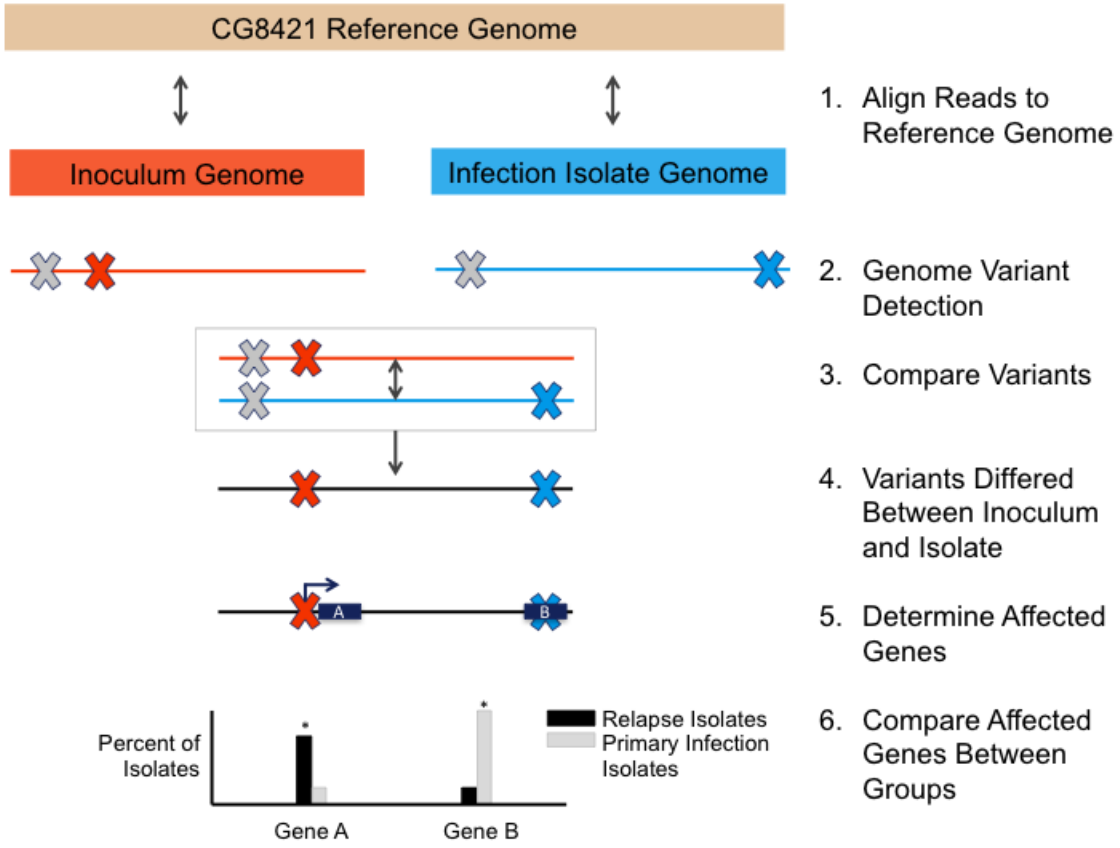



Illustration 2.1 Schematic of Genome Variant Analysis

The genomes of the *C. jejuni* CG8421 inoculum and infection isolate population genomes were compared to the published reference genome. Genome variants that occurred in at least 1 percent of any sample population are listed in Sup Table 1C. However, to identify influential variants we used a 25 percent frequency cutoff as described in the methods and paper text, and those variants are listed in Sup. Table 4. The variant calls between the inoculum genome and the infection isolate populations where then compared, and variant calls that differed between the inoculum and isolate populations are considered genomic variants selected for *in vivo*. Using the location of each variant in the annotated genome we determined the genes (or gene promoters) affected by these variants, and then used Fisher's Exact Test with FDR-corrected p-values to determine if particular genes were more likely to genetically vary in *in vivo* between volunteer isolate groups.

These infection isolate populations represent bacterial colonies harvested directly from primary isolation plates and are pooled together per sample with no further outgrowth, and is consistent with previous work that validated *C. jeuni* isolate sequencing

to study genetic variation *in vivo*[26]. We achieved exceptional coverage of the infection population genomes across all samples, with an average genome sequencing coverage >1000x and 97 percent of genes had at least 25x coverage at every nucleotide position (Fig. 2.3).

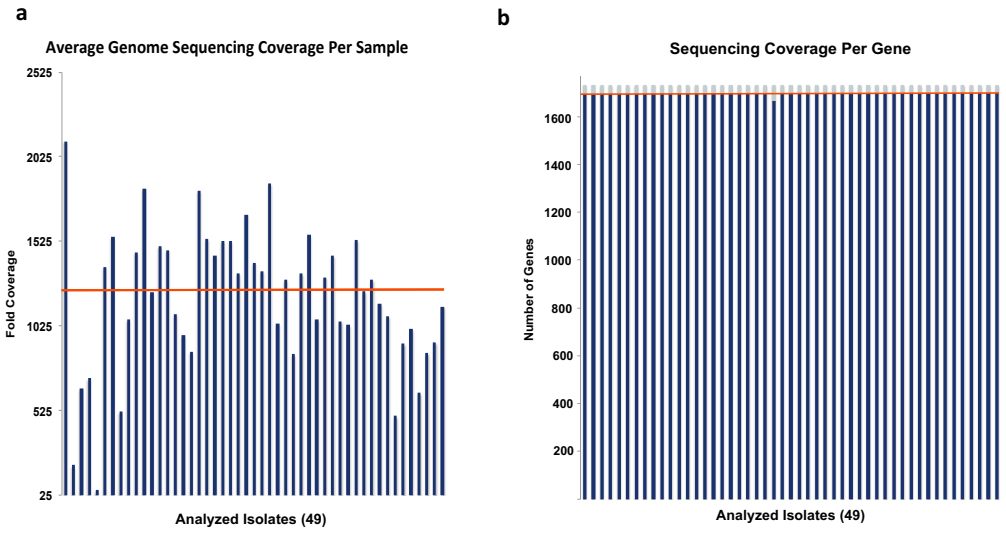



Figure 2.3 Infectious Isolate Populations Genome Sequencing Coverage

**a,** Sequencing depth across all 49 infection isolate populations. The average fold coverage across samples was 1192x (orange line). **b**, Coverage of open reading frames across samples. On average (orange line) 98 percent of annotated genes had at least 25 fold coverage at every base across samples. Grey bar represents the number of genes where at least one nucleotide had <25x coverage. Blue bar represents the number of genes where every nucleotide in the gene had >50% of mean coverage (1192x mean coverage across all samples).

Due to the complexity of phase variable populations, different subjective methods have been used to determine noteworthy changes in phase variant frequencies over time[26,54,151]. We first called any variant that occurred in at least 1 percent of any sample population, which resulted in over 600 variants identified across all samples. To focus on variants that had both a large change in frequency during infection and represented a

large portion of the population, we called variants in a sample if they occurred in at least 25 percent of the population. This enriched for variants that had, on average, a large statistically significant fold change in frequency between the inoculum and an infection isolate populations (~12 fold) while also representing a majority of the infection population sample (~60 percent) (Fig. 2.4).

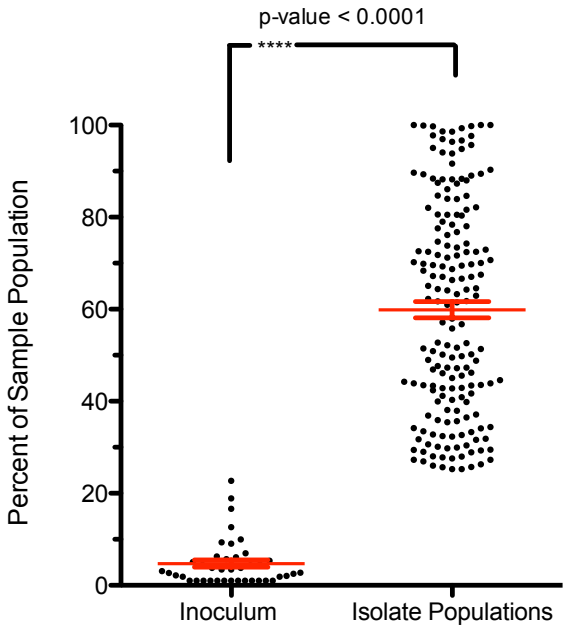



Figure 2.4 Average Frequency of Variants Called in Isolate Populations

Genomic variants were initially called if they occurred in at least 1 percent of a sample population (49 isolate samples and 1 inoculum), resulting in over 600 variants detected across all samples, and that data is found in Sup. Table 3. To focus on variants that had both a large change in frequency during infection and represented a large portion of the population, we used a 25 percent variant calling cutoff. This raised cutoff enriched for variants that had, on average, a large fold change in frequency between the inoculum and an infection isolate populations (~12 fold) while also representing a majority of the infection population (~60%). Detailed information about each variant call, per sample, is provided in Sup .Table 5 and is represented here by a dot plot. An unpaired two-tailed T-test was used to confirm a significant difference in the mean frequency between the pre (inoculum) and post infection (Isolate Population) variant frequency calls. Bars represent the mean variant frequency in the sample(s) and standard error of the mean.

This enrichment analysis resulted in 48 unique genomic variants with major changes in frequency in at least one sample, including small nucleotide polymorphisms

(SNPs), phase variations (PV), and multinucleotide variants (MNV). 47 of the variants clearly affected 30 annotated coding sequences and/or promoters and are noted per isolate population in Figure 1.2b. Slip-strand mutations in poly G/C nucleotide tracts are the dominant phase variable mechanisms in *C. jejuni*[26], and 19 of 28 tracts in the CG8421 genome varied in at least one infection isolate population (Fig. 2.5). The inoculum population had 12 genetic variants that differed from the reference genome, 4 of which remained intact in all infection isolate populations.

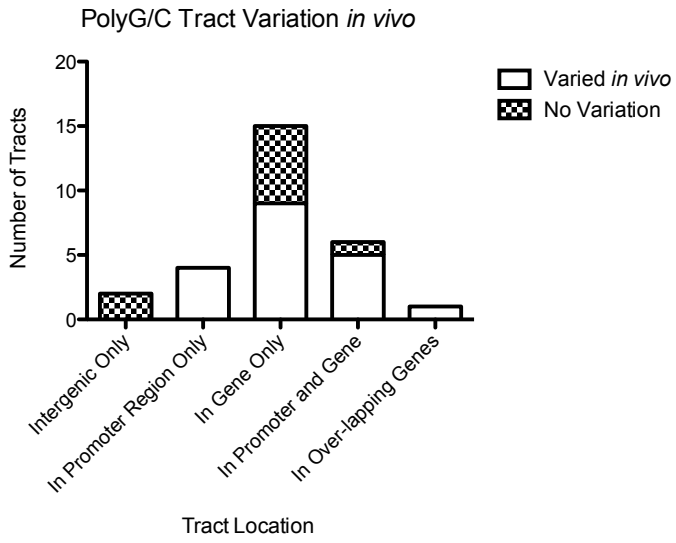



Figure 2.5 Poly G/C Tract Locations and Variations in Vivo

The location of all 28 poly G/C tracts in the CG8421 genome are presented by their relationship to annotated open reading frames in a stacked bar graph. The number of tracts at each location type that varied in at least one sample population are noted.

Three genetic variants were selected for in all 49 infection isolate populations (Figure 1.2b). One of these completely conserved variants is a PV that occurs in the in last 10 percent of the *fedA* gene open reading frame and in the -10 promoter region of an overlapping downstream gene of unknown function. *fedA* is putative hemerythrin that contributes to *C. jejuni* colonization in chicks and is coregulated with flagellar gene expression[36]. A second completely conserved variant is a SNP in the house keeping gene

adenine phosphoribosyltransferase (*apt*) which salvages adenine to make adenosine monophosphate. The inoculum population contained an *apt* SNP that is selected for during osmotic stress[54], however the wild type *apt* sequence was selected for *in vivo*. The final completely conserved variant is a 9 nucleotide in-frame deletion in the CG8421 homolog of the DNA response regulator CJJ81176_1483 that was selected against *in vivo*. Selection for the wild type version CJJ81176_1483 *in vivo* was confirmed by structural genomic variant analysis. CJJ81176_1483 is part of a newly identified two-component system that regulates a gluconate dehydrogenase complex important for chick colonization[165]. These data represent 3 distinct genetic variations (PV, SNP, and MNV) in genes with diverse roles (chicken colonization, osmotic stress/house keeping, two-component system) and demonstrate the genetic adaptability of *C. jejuni in vivo*.

### 2.2.3 Gene Variants Associated with Treatment or Human Disease State

Rifaximin prophylactic treatment did not impact the rate of campylobacteriosis during the clinical trial. Nevertheless, we hypothesized the antibiotic treatment environment in the host may have selected for unique genomic variants in these volunteers. However, we found no variants were more likely to be called in these volunteers when compared to control volunteers using Fisher's Exact Test. We concluded rifaximin did not exert a noticeable pressure on *C. jejuni in vivo*. We also hypothesized some variants may be more virulent, and therefore would be more likely to be found in volunteers who experienced severe disease symptoms. However, a Fisher Exact Test showed no variants were associated with isolates from volunteers experiencing severe disease compared to typical symptomatic disease isolate populations. This indicates host factors, more so than genetic drift of more virulent genotypes, may influence the severity of disease symptoms.

Despite being treated with azithromycin and ciprofloxacin and repeatedly producing feces negative for *C. jejuni* isolates at the end of the trial, 5 volunteers returned to the clinic with at least one bout of recrudescent infection (Fig. 2.2a). This rate of recrudescence is higher than previous trials likely due to enhanced culture techniques, such as not refrigerating stool before culturing, that were employed specifically for increased vigilance of recrudescence[61]. Recrudescent infection isolates were confirmed to be the inoculum strain and, in all cases, were sensitive to azithromycin and ciprofloxacin. Remarkably, relapse infection isolate populations had twice as many genomic variants as primary infection isolates (Fig. 2.6a). We hypothesized the increased variant count would correlate with increased time *in vivo*. To test this, we compared the number of variants called per sample on every day of infection. Interestingly, there was not an increase in the number of variants called over time during the primary infection (Fig. 2.6b) and a Fisher exact test showed no particular variants are associated with early or late primary infection periods. These data suggest there is an immediate and consistent selection pressure during primary infection, and that either a secondary selection event or additional time within the host results in increased genetic variation of relapse isolates.

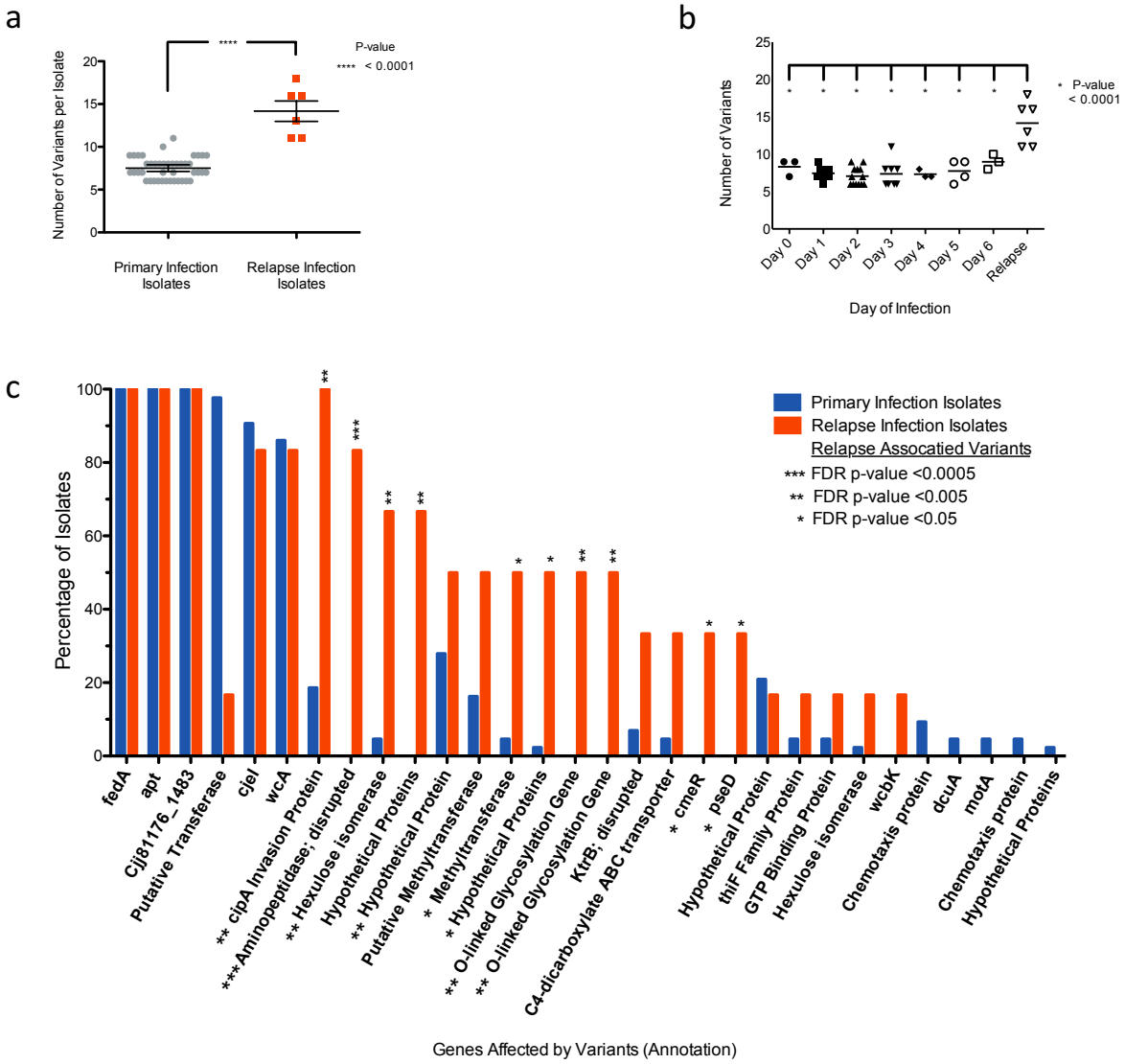



Figure 2.6 Particular Genome Variants are Associated with *C. jejuni* Recrudescent Infection Isolate Populations

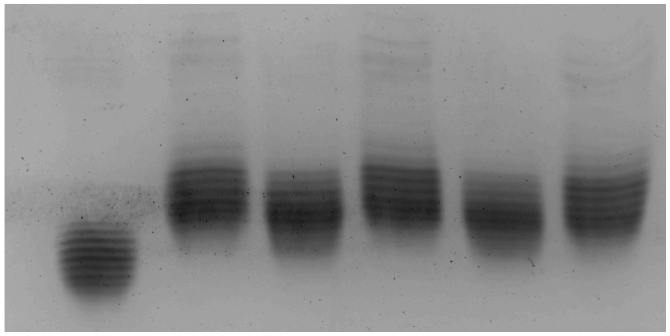
Caption for Figure 2.6 begins on following page.

Caption for Figure 2.6:
Primary infection isolates (days 0-6) represent 43 isolate samples while relapse infection isolate populations represent 6 samples. **a,** Bulk genome variation between primary and relapse infection isolates. On average, relapse infection isolates have twice as many genomic mutations compared to primary infection isolates (Unpaired two tailed T-test with bars representing the mean and standard error of the mean). **b**, Variant accumulation over time. Variants are selected for immediately in human infections (day 0 is the day of inoculation); however, they do not increase over time during the primary infection (One-way ANOVA, Tukey's multiple comparison test). **c**, Genome variants that are associated with relapse infection. After the genome variants were determined for all samples (Fig 2b), a Fisher's Exact test was used to determine if variation in particular genes is associated with relapse infection isolate populations. Variation of 10 genes was statistically associated with relapse infection populations and is noted with a False-Discovery Rate adjusted (FDR) one-sided p-value of $< 0.05$. The y-axis represents the percent of samples that had a variant in the gene noted on the x-axis. More detailed information is noted in Sup. Table 4.

There are a considerable number of genetic variants that are associated with relapse infection isolates as determined by Fisher Exact Test (false-discovery rate corrected p-value <0.05) (Fig. 2.6c, Sup. Table 10). Of the 10 gene variants associated with relapse infection only 3 are characterized, named genes. The most striking is a phase variation in *cipA* (cell invasion protein A), which turned the gene "on" in every relapse infection isolate population, but does not appear to be important for establishing a primary infection. *cipA* was identified in three independent studies to contribute to *C. jejuni* invasion of host cells *in vitro*[34,166]*;* however, this gene has not been shown to play a role *in vivo* until this work in a human challenge model. Mechanisms of *C. jejuni* cellular invasion are not well defined and no biochemical functions of CipA have been previously identified. Although, BLAST search revealed *cipA* harbors a domain of unknown function (DUF2972) that is also found genes annotated as glycosyl transferases. To determine the role of CipA during human infection we therefore considered possible glycosylation targets. We find the "on" variant of *cipA* that was selected for *in vivo*

results in modification of the flagella (Figure 1.8a). This suggests *cipA* likely contributes to cellular invasion and persistence in humans via flagella modification.

**a**




**b**

| **Fisher's exact test** ** P-value 0.0045 | Disease state associated (either primary or relapse infection association) | No Associations | Total |
|---|---|---|---|
| Flagellin modification genes | 6 | 1 | 7 |
| All other *in vivo* variable genes | 5 | 18 | 23 |
| Total | 11 | 19 | 30 |

**c**

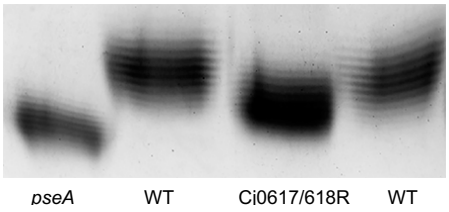



Figure 2.7 Genetic Variation in Flagella Modification Genes is Strongly Associated with the Human Disease State

Caption for Figure 2.7 begins on the following page.

Caption for Figure 2.7:
**a,** A functional *cipA* gene results in modification of the flagella as shown by an isoelectric focusing gel (pH 3-5) of purified flagellins from *C. jejuni* strain 81-176. The *cipA* gene was phase varied "off" in the human inoculum, but was selected to be turned "on" in every human relapse infection isolate population. Therefore, a functional *cipA* gene is strongly associated with relapsing human infections. We therefore aimed to identify the biochemical role of *cipA* in *Campylobacter*. *cipA* contains a domain of unknown function found in putative glycosyl transferase genes as predicted by BLAST search. We therefore sought to identify possible targets of *cipA* activity in *Campylobacter*. We investigated potential flagella modification by preforming an isoelectric focusing gel mobility shift assay on purified flagella from *cipA* "on" and "off" genetic backgrounds. Flagellin modification activity changes the electrical charge of the flagellin and therefore results in a mobility shift in the delicate isoelectric focusing gel. Therefore, in lieu of standard ladders, purified flagella from characterized intact and mutant flagella modification strains are used as simultaneous markers and reference controls to interpret modification-induced shifts on IEF gels. Genetic manipulation cannot be performed in strain CG8421 so strain 81-176 was used for these analyses. Lanes contain flagellin from: *pseA*, a mutant lacking the acetamidino form of pseudaminic acid, serving as a control demonstrating a more negatively charged modification state compared to wildtype; WT, wildtype 81-176 flagellin in which *pseA* is intact, however *cipA* is "off" similar to the human inoculum; c*ipA*R, a mutant in which the phase variable homopolymeric G tract in *cipA* was repaired to produce a full-length *cipA* open reading frame in an otherwise WT background. Indeed, a mobility shift of flagella from the *cipA* "on" repaired strain indicates a more negative charge and therefore *cipA* contributes to flagellin modification. Samples were repeated in multiple lanes for clarity of the phenotype and this gel is representative of at least 5 mobility shift experiments. **b,** Contingency table of variable genes and their association with either primary or relapse infections (disease state). A Fisher's exact test shows variation in flagella modification genes (including *O*-link glycosylation locus genes as described in Sup. Table 4) as a group is strongly associated with either primary or relapse infection isolate populations (one-sided p-value). The 8 variable genes related to flagellin glycosylation[1] are Cj11168 homologs *Cj0617/618*, *Cj1295*, *Cj1306c*, *Cj1325*, *Cj1342*, *Cj1333*, *Cj1321* (*CJ8421_RS06560*), and *cipA* as identified in Sup. Figure 6a. **c,** Isoelectric focusing gel of flagellins from *C. jejuni* strain 81-176 and 81-176 mutants. *Cj0617* and *Cj0618* are phase variable genes that can be translated as a unified open reading frame (*Cj0617/618*). Although *Cj0617* is the representative gene of the 617 family of *O*-linked glycosylation genes[2], and *Cj0618* contributes to motility[3], *Cj0617* and Cj0618 are the only genes included in Sup. Fig. 6b, besides *cipA*, that are not located in the *O*-linked glycosylation locus and there are no reports showing the affect of *Cj0617/618* on flagellin modification to our knowledge.

Caption for Figure 2.7 continues on the following page.

Figure 2.7 caption continued…We therefore sought to verify that *Cj0617/618* contributes to flagella modification by preforming an isoelectric focusing gel mobility shift assay on purified flagella. Flagellin modification activity changes the electrical charge of the flagellin and therefore results in a mobility shift in the isoelectric focusing gel. Therefore, in lieu of standard ladders, purified flagella from characterized intact and mutant flagella modification strains are used as simultaneous markers and reference controls to interpret modification-induced shifts on IEF gels. Genetic manipulation cannot be performed in strain CG8421 so strain 81-176 was used for these analyses. Lanes contain flagellin from: *pseA*, a mutant lacking the acetamidino form of pseudaminic acid, serving as a control demonstrating a more negatively charged modification state compared to wildtype; WT, wildtype 81-176 flagellin in which *pseA* is intact, however Cj0617/618 is not translated as a continuous open reading frame; *Cj0617/618*R, a mutant in which the phase variable homopolymeric G tract was repaired to produce a continuous open reading frame of genes *Cj0617* and *Cj0618* in an otherwise WT background. Indeed, a mobility shift of flagella from the *Cj0617/618* repaired strain indicates a more negative charge and therefore *Cj0617/618* contributes to a flagellin modification phenotype. This mobility shift is representative of at least 5 mobility shift experiments. *Cj0617/618* likely affects pseudaminic acid modifications on flagellin as strain 81-176 lacks the alternative legionaminic acid pathway.

Another notable gene associated with relapse infection is *cmeR*, a pleotropic bile-sensing transcriptional regulator that is required for chicken colonization[130]. CmeR functions as a dimer to bind DNA and repress gene expression until *C. jejuni* is exposed to bile in the intestinal environment[134]. CmeR repression is relieved upon bile acid binding, which induces expression of the CmeR regulon including the multi-drug efflux pump *cmeABC*[167]. Two unique SNPs that each resulted in truncation of the *cmeR* open reading frame were selected for independently in relapse infection isolates and are in non-phase variable loci. Similar truncations have been shown to prevent dimerization of CmeR and therefore prevent CmeR repression regardless of environmental bile concentrations[168]. *cmeR* was intact in every primary infection isolate, and likewise, functional CmeR protein is important for colonization of chickens. However, the independent and functionally conserved relapse variants suggest a *cmeR* "off" mutation can be advantageous in persisting human infections. This may be due to a resulting constitutive expression of the CmeABC efflux pump that increases resistance to bile

acids in the host[133], although losing the ability to sense bile is an unexpected *in vivo* phenotype for an intestinal pathogen.

*pseD* is the final named gene variant that is associated with relapse infection. PseD attaches pseudaminic acid to flagellin which contributes to virulence-associated autoagglutination of the bacteria[169]. *pseD* was considered "on" in every primary infection sample but turned "off" in a one third of relapse infection samples. The remaining 7 relapse associated gene variants included hypothetical genes, a disrupted aminopeptidase gene, a methyltransferase gene, and two additional *O*-linked glycosylation locus genes. Interestingly, an annotated methyltransferase and multiple *O*-linked glycosylation locus genes were identified to be under positive selection in a study of *Burkholderia* infection of cystic fibrosis patients[6], although these genes share low sequence homology with genes in CG8421[6].

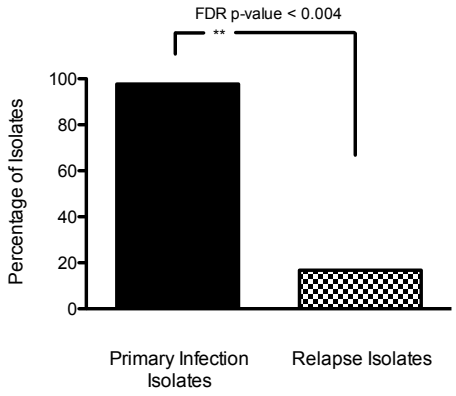



Figure 2.8 Genome Variant Associated Only With Primary Infection Isolate Populations

A phase variant in the promoter region of the uncharacterized putative transferase gene CJ8421_RS06560, present in the inoculum, was resoundingly selected against during primary infection of humans. The percentage of isolates shown denotes the frequency that the wildtype (reference sequence) variant was called in those samples. Fisher's exact test identified this variant as statistically significantly associated with primary infection isolates (43 samples) compared to relapse infection isolates (6 samples) with a false-discovery rate (FDR) correct p-value.

One variant is statistically significantly associated with primary infection isolates and not relapse infection isolates (Fig. 2.8). This is a phase variant in the promoter region of an uncharacterized putative transferase gene, *CJ8421_RS06560* (homolog of *Cj1321* in strain NCTC_11168), located in the flagellin glycosylation locus[170] and was strongly selected against in primary infection isolates, but not in relapse infection isolates. In total, 7 of the 8 flagella modification genes identified in our analysis are associated with either primary or relapse infection isolates, indicating the significance and variability of flagellar modification during human infections (Fig. 2.7b). Finally, Figure 2.9a depicts a SNP-tree analysis to visually represent the differences in genomic variations amongst primary and relapse infection isolates. The closer two samples are on the tree, the more similar their genomic variant profiles are to one another. Primary infection isolates from the same volunteer are more similar to each other than they are to the isolates of other volunteers. However, the relapse infection isolates from different volunteers are more similar to each other than they are to the founding primary infection isolates from the same volunteer. Overall, the relapse-associated variants demonstrate the profound genetic differences that occur in persisting bacteria.

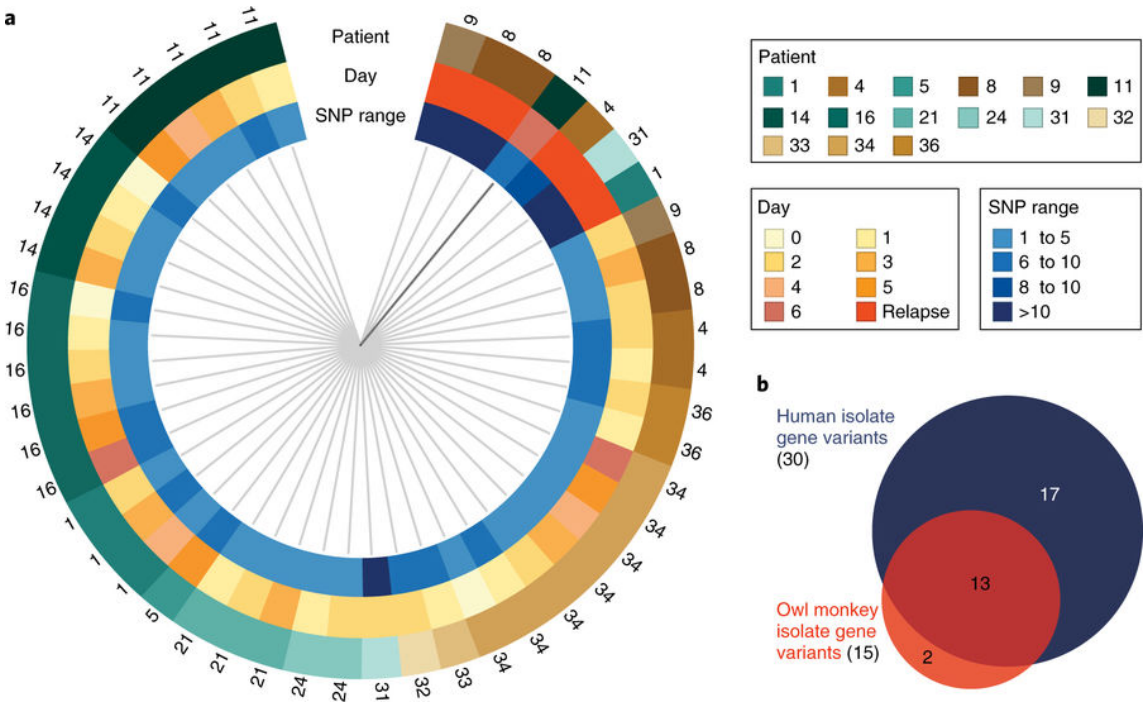



Figure 2.9 Validating *C. jejuni* Genomic Variant Selection in non-Human Primates

**a**, A SNP tree representing how similar human infection isolate populations are to one another. Each wedge represents one sequenced isolate population and the closer two samples are to each other the more similar their genomic variations. Volunteer identification numbers per sample are noted on the outside of the ring, and the day of infection and the number of SNPs detected are denoted by the inner rings. Note that primary infection isolates from the same volunteer are more similar to each other than they are to isolates from other volunteers. However, relapse infection isolates (bright orange) are more similar to each other despite the volunteer they come from. **b,** Conserved gene variant selection across primates. A second, different *C. jejuni* CG8421 inoculum was used to infect owl monkeys, and the genomes of 12 diarrheal primary infection isolate populations from 5 monkeys were sequenced (Table 1.1). The Venn diagram shows the *C. jejuni* genome variants selected for in owl monkey infection closely resemble those of human infection. This supports the human data and suggests non-human primates harbor a similar fitness selection environment found in humans. The smaller sample size and no relapse infections contributed to fewer total variants detected in the monkeys.

### 2.2.4 Genome Variant Selection in *Aotus* Primates

The variants selected for during human infection replicated strongly across volunteer isolate populations. This powerfully suggests a conserved selective pressure and adaptive response *in vivo*. However, over 30 coding sequences were implicated in adaptation, and understanding the fitness contribution of variant combinations within the population is challenging. To validate our data and respect the complexity of population-level adaptation, we performed a similar analysis in a symptomatic non-human primate infection model. We hypothesized adaptations similar to those in human infection would be selected for in a New World monkey, *Aotus nancymaae*. An independently prepared CG8421 liquid inoculum was used to infect *Aotus nancymaae* and we compared this inoculum genome to 12 diarrheal isolate populations taken over 8 days of infection from 5 monkeys (Table 2.1).

| | | | Days After Inoculation | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Animal Number** | Animal Sex | Animal Age (months) | **2** | **3** | **4** | **5** | **7** | **8** | **9** | **10** |
| **1** | Male | 13 | | | | | | | | |
| **2** | Female | 14 | | | | | | | | |
| **3** | Female | 13 | | | | | | | | |
| **4** | Male | 17 | | | | | | | | |
| **5** | Female | 14 | | | | | | | | |

**Legend**

| | |
|---|---|
| | Non-diarrhea sample (stool grade < 3) or insufficient C. jejuni genome coverage for analysis and/or sample not used in this study |
| | Diarrhea sample (stool grade >3), isolate population sequenced, sufficient C. jejuni genome coverage, sample used in this study |

Table 2.1: Non-human primate infection and samples used.

Despite differences between the human and *Aotus* inoculums, we found the genetic adaptations selected for in *Aotus* closely resemble those in humans (Fig. 2.9b). The 3 completely conserved variants that were selected for in human infection were already present as a majority in the *Aotus* inoculum, and congruently, were maintained during infection. Two genes were uniquely varied in *Aotus* infection, including a SNV in the promoter region of *CJJ81176_1215*, a gene encoding an uncharacterized, putative NLPA-family lipoprotein. Interestingly, the second variant is a E70K amino acid substitution in *chuA*, which is also the most upregulated gene during human infection. Therefore, *chuA* can undergo both transcriptional and genetic adaptations during primate infection. Few of the genetically variable genes were also differentially regulated *in vivo*, yet this is consistent with analyses done *in vitro*[171] and *in vivo*[172] that show genes that contribute to fitness are rarely also differentially regulated.

## 2.3 DISCUSSION OF *C. JEJUNI IN VIVO* ADAPTATION

Understanding the bacterial pathogen response to the human host revealed genetic mechanisms of success during both establishment of infection and persistent human colonization. Notably, we determined the *cipA* gene had the strongest association with recrudescent, persistent infections, and we went on to identify its role in *C. jejuni* flagellar modification. 9 other variable genes are part of the conserved relapse infection genotype. We suspect this genetic background may enable bacteria to reside in a protected niche within the host during primary infection where they withstand being shed from the host, resist antibiotic treatment and the host immune response, and result in recrudescent infection. This genotype would then be the dominant recrudescent population, and a similar pattern was seen in this study. Future therapy design should consider the "on/off" status of a therapeutic target throughout primary and recrudescent infections, as well as appreciate target expression alterations due genetic variation of promoter regions or transcription factors such as CmeR.

Tracking pathogen evolution *in vitro* requires studying multiple cultures grown in parallel, and variants that arise consistently across replicates are considered adaptive. Studying pathogen evolution *in vivo* can reveal important virulence determinants, although it is difficult to study well-controlled human infections in parallel. Here we used a *C. jejuni* clinical trial to examine bacterial adaptation during human infection and found many genomic variants are well conserved across infected volunteers. We also found human infection exemplifies rapid bacterial adaptation, as variants are selected for *in vivo* on the day of inoculation. This study characterizes pathogen genetic and transcriptional adaptations across hosts and lifestyles, and similar collaborative approaches in future clinical studies may increase the possibilities for therapeutic design.

# Chapter 3: Enterotoxigenic *E. coli* Virulence Gene Regulation in Human Infections

## 3.1 INTRODUCTION

Enterotoxigenic *Escherichia coli* (ETEC) is a global gram-negative diarrheal pathogen of the small intestine that infects up to 200 million people worldwide each year[3]. ETEC infections result in hundreds of thousands of deaths annually and rely on a trio of classical ETEC virulence factors to cause disease[3,80,4]. The prototypical ETEC strain H10407 harbors heat-labile and heat-stable enterotoxins along with the colonization factor antigen (CFA) adhesin[85]. The production of CFA is controlled by the transcriptional regulator CfaD, which activates the expression of the *cfaABCE* operon on virulence plasmid p948[173]. The outer membrane usher protein CfaC and periplasmic chaperon CfaA assemble the extracellular CfaB pillus with the CfaE adhesive tip[87,95]. Robust CFA adhesin production occurs in CFA media, a namesake standard ETEC complex growth media.

Alongside the CFA adhesin, H10407 also harbors the heat labile toxin and two homologs of the heat stable toxin denoted as *sta1* and *sta2*[80,110,174]. The enzymatic activity of these enterotoxins results in water efflux from host intestinal cells which produces profuse diarrhea[175,176]. *sta1* and *sta2* are found on plasmids p666 and p948, respectively. Despite their conserved functional role, *sta*1 and *sta2* have different promoters and their expression can be inversely regulated[129]. Recent work has shown the cyclic AMP receptor protein (CRP), in response to changes in extracellular glucose concentrations, can oppositely regulate *sta1* and *sta2* promoter::*lacZ* fusion reporter constructs in *E. coli* K-12[129]. The significance of differential regulation between two heat stable toxin promoters remains unclear, although it demonstrates how a single environmental factor can have a complex effect on virulence factor expression in ETEC. Heat labile toxin is a

heterohexameric, A-B subunit enterotoxin encoded by the *eltAB* operon on virulence plasmid p666[177]. It is functionally and structurally similar to the diarrhea inducing cholera toxin produced by *Vibrio cholerae*[177,178].

Although ETEC H10407 contains all of these classical virulence factors, only one enterotoxin is required for symptomatic disease[86]. ETEC also contains a large group of non-classical[179] virulence factors that have been shown to contribute either directly or indirectly to virulence-related phenotypes in *E. coli*. Notably, genes *etpABC*[180] and *tibAB*[181] were originally described in ETEC, and have been shown to enable host cell adhesion *in vivo* and host cell adhesion/invasion *in vitro*, respectively.

Decades of research on classical virulence factor gene regulation in ETEC has demonstrated how over 20 environmental conditions influence expression of individual virulence genes, including pH, bile, bicarbonate, osmolarity, and glucose[129,173,182]. Understanding how ETEC responds to these signals in the mammalian intestinal tract for effective colonization and transmission is challenging, particularly considering just one of these factors can alter expression of virulence factors in opposing ways. For example, it was recently shown glucose can oppositely regulate heat stable and heat labile toxin promoter::*lacZ* fusions[129], and there are scarce reports of all virulence factors regulated in a similar manner simultaneously.

Successful intestinal colonization requires appropriate bacterial transcriptional adaptation to the host environment. We therefore examined the ETEC transcriptome in human infection samples to understand how ETEC responds to the host *in vivo*. The host-specific response has important implications for therapy design as it can reveal that high priority drug and vaccine targets may have unexpected expression patterns *in vivo*[9]. Here we used samples from a controlled human infection model to examine the global

transcriptome of ETEC H10407 directly in human infection samples (ClinicalTrials.gov Identifier NCT01922856).

The human challenge was developed to test the efficacy of a vaccine immunization against CfaE and heat labile toxin to prevent diarrheal episodes in humans and will be detailed elsewhere. We performed RNA-sequencing directly on the stool samples of 5 infected volunteers who were not given vaccine treatment. Many studies have investigated the expression of individual ETEC genes in infected stool samples; however, variation between wild type isolates can make interpretation of these data challenging. In this study, each volunteer was infected simultaneously from the same H10407 inoculum. This allowed us to study biological replicates of human infection without confounding differences between strains or inoculums, and further, we performed a global analysis of ETEC gene expression directly in the infected samples using next-generation high throughput RNA-sequencing.

We were surprised to find the heat-labile toxin genes and CFA adhesin expression were down regulated in infected stool samples compared to laboratory growth in CFA media. After considering the stool samples represented the anaerobic large intestinal environment, and that little work has been done to characterize ETEC in low-oxygen conditions, we examined the ETEC transcriptome during growth in microaerobic and anaerobic atmospheres. Besides the anaerobic large intestinal environment, anaerobic conditions in the proximal small intestine are fortified by bacterial overgrowth[183] and the distal small intestinal lumen naturally has near zero levels of oxygen[184]. Remarkably, we discovered anaerobic *in vitro* growth fully replicated the virulence gene expression pattern seen in the stool samples, indicating environmental oxygen played an important role in understanding virulence factor expression *in vivo*. Anaerobic growth *in vitro* results in a significant decrease in expression of *cfaD*, the *cfa* operon, and the *eltAB*

operon, and an increase in *sta2* expression when compared to aerobic growth. We went on to determine the effect of oxygen on virulence gene expression is dependent on the oxygen-sensitive transcriptional regulator FNR[185,186] (Fumarate Nitrate reduction Regulator). Deletion of *fnr* in H10407 resulted in a significant increase in expression of all classical virulence factor genes, including both *sta1* and *sta2*, indicating FNR activity results in a global repression of virulence gene expression in ETEC.

We also found that H10407 produces significantly more biofilm when grown anaerobically compared to aerobic growth, and that a *Δfnr* H10407 strain produces significantly more biofilms than WT under anaerobic conditions. In afflicted areas, ETEC biofilm formation on abiotic surfaces is correlated with seasonal epidemics[84], which suggests FNR-dependent oxygen-sensing may contribute to human ETEC infection both inside and outside of the host.

Overall, by examining the transcriptome of ETEC in infection samples from a controlled human infection model, this study identified an environmental factor and corresponding transcriptional regulator that influences global virulence and biofilm regulation. This work suggests the low oxygen environment of the intestinal lumen likely represses ETEC virulence gene expression *in vivo* - particularly when bacterial overgrowth fortifies anaerobic conditions in the proximal small intestine[183,187], in the distal small intestine where luminal oxygen levels are near zero[184], and in the anaerobic large intestine during transmission[184]. However, oxygen seepage from intestinal epithelial cells is sensed by *E. coli* via FNR *in vivo*, and the highest oxygen concentrations exist nearest the epithelium[188,189]. Therefore, ETEC virulence gene expression is likely repressed in the low-oxygen lumen and is relieved when the oxygen-intolerant FNR is inactivated near the epithelium.

## 3.2 RESULTS

### 3.2.1 Controlled Human Infection Samples and RNA-sequencing

Transcriptional adaptation to the host environment is essential for intestinal pathogens to colonize and cause disease. To understand how ETEC adapts to the human host we characterized the transcriptome of ETEC H10407 directly in infected diarrhea samples taken from five different volunteers in a controlled human infection model. The same inoculum was used to infect volunteers at 1.2x10^7 colony forming units each. The samples used in this study came from volunteers who were not prophylactically treated or vaccinated. These volunteers were not prescribed a fixed diet or specific rehydration therapy during their infection.

RNA extracted from stool mostly represents RNA derived from host cells, the native microbiome, and food waste. To specifically identify ETEC transcripts in the dataset, reads were trimmed to 21 bases long and mapped with 100% exact homology to the H10407 reference genome as previously described for similarly complex samples[9]. To be sure commensal *E. coli* were not influencing our RNA-sequencing data we measured the ratio of mapped ETEC chromosomal transcripts to transcripts that mapped to virulence plasmids. If a volunteer had a prominent commensal *E. coli* colonization, which would likely produce transcripts that are highly homologous to ETEC chromosomal transcripts, we expected ETEC chromosomal transcripts would be over-represented compared to ETEC-specific virulence plasmid transcripts. We did not find this to be the case for any of our sequenced samples.

### 3.2.2 Environmental Oxygen Explains *in vivo* Virulence Gene Expression

We determined the ETEC transcriptome directly in 5 infected diarrhea samples, each from different volunteers, which together as a sample group represent our *in vivo* transcriptome. 784 genes were upregulated and 637 genes were downregulated *in vivo* compared to aerobic mid-log phase growth in CFA broth (Figure 3.1a).

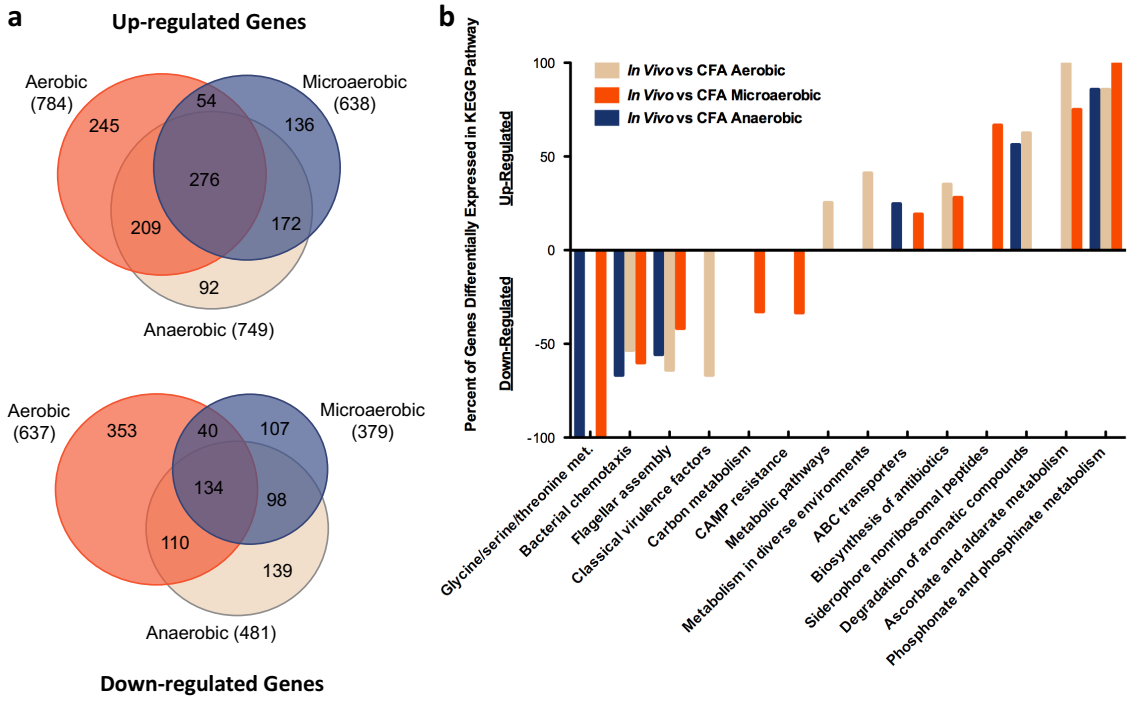



Figure 3.1 ETEC gene regulation *in vivo*

**a.** The number of upregulated or downregulated genes in infected stool compared to mid-log phase *in vitro* CFA broth growth in the indicated atmospheres. Microaerobic atmosphere was 85 % nitrogen, 10 % $CO_2$, 5 % $O_2$. Anaerobic atmosphere was 95 % nitrogen, 5 % hydrogen, 0 % oxygen. **b.** Kegg orthology pathways statistically significantly ($p < 0.05$) represented in genes differentially regulated *in vivo* (fold change >|5| FDR $p < 0.05$) using a hypergeometric test. Classical virulence factors were manually curated into our KEGG pathway list to be included in the analysis. *In vivo* samples represent 5 infected diarrhea samples from different human volunteers and *in vitro* conditions each represent three independent biological replicates.

Given the close association with the host we hypothesized classical virulence factor genes would be upregulated in the stool vs. *in vitro* growth. Surprisingly, classical virulence factors were largely downregulated in stool compared to standard laboratory growth, including *cfaD*, the *cfa* operon, and *eltAB* (Figure 3.1b, Figure 3.2).

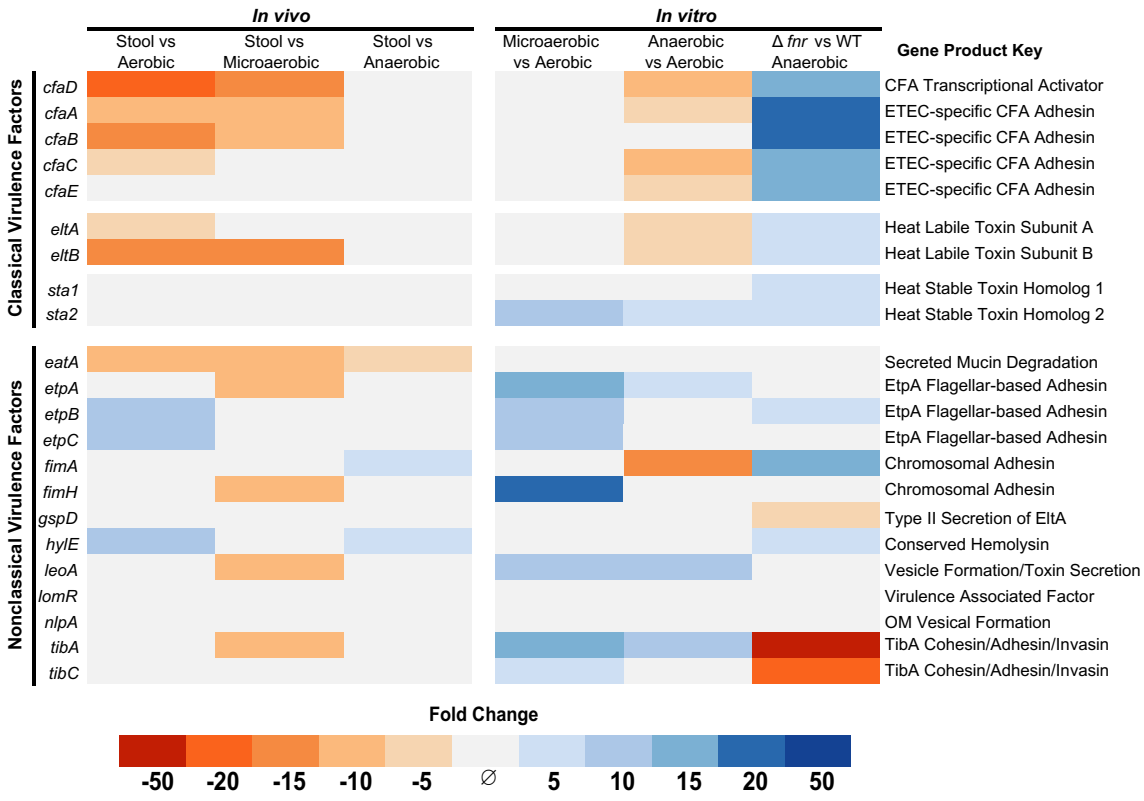



Figure 3.2 Oxygen and FNR-Dependent Virulence Gene Expression in ETEC H10407

Differentially regulated (Fold change >|3| fold, false-discovery rate corrected p-value <0.05) ETEC virulence genes *in vivo* compared to *in vitro* growths in different atmospheric oxygen concentrations. *In vitro* anaerobic growth replicated the *in vivo* expression pattern of classical ETEC virulence factors. *In vivo* samples represent 5 infected diarrhea samples from different human volunteers and *in vitro* conditions each represent a group of three independent biological replicates.

Stool is frequently used as a proxy for understanding the state of the intestinal environment, and a hallmark of the intestinal tract is low availability of oxygen[189]. The difference between environmental atmospheres was a major discrepancy between our aerobically grown laboratory comparison and ETEC in infected stool. Despite the *in vivo* relevance of growth in limited oxygen atmospheres, we found little published research describing ETEC in such conditions.

To address this, we grew biological triplicates of ETEC H10407 in microaerobic (85% N, 10% $CO_2$, 5% $O_2$) and anaerobic (95% N, 5% H, 0% $O_2$) atmospheres to mid-log phase with shaking (Figure 3.3) and compared these low-oxygen transcriptomes to the *in vivo* stool transcriptomes (Figure 3.1, Figure 3.2).

Remarkably, anaerobic growth in CFA broth completely replicated the virulence factor expression profile seen *in vivo*, with no statistically significant difference in expression of any classical virulence factor between the two conditions (Figure 3.2). This suggests understanding ETEC's response to environmental oxygen is important for understanding ETEC virulence gene regulation *in vivo*. Notably, the expression of both heat stable toxin genes did not differ *in vivo* compared to any environmental oxygen concentration. This lack of statistical significance likely reflects variation in the expression of *sta* genes between the complex patient samples (Figure 3.4).

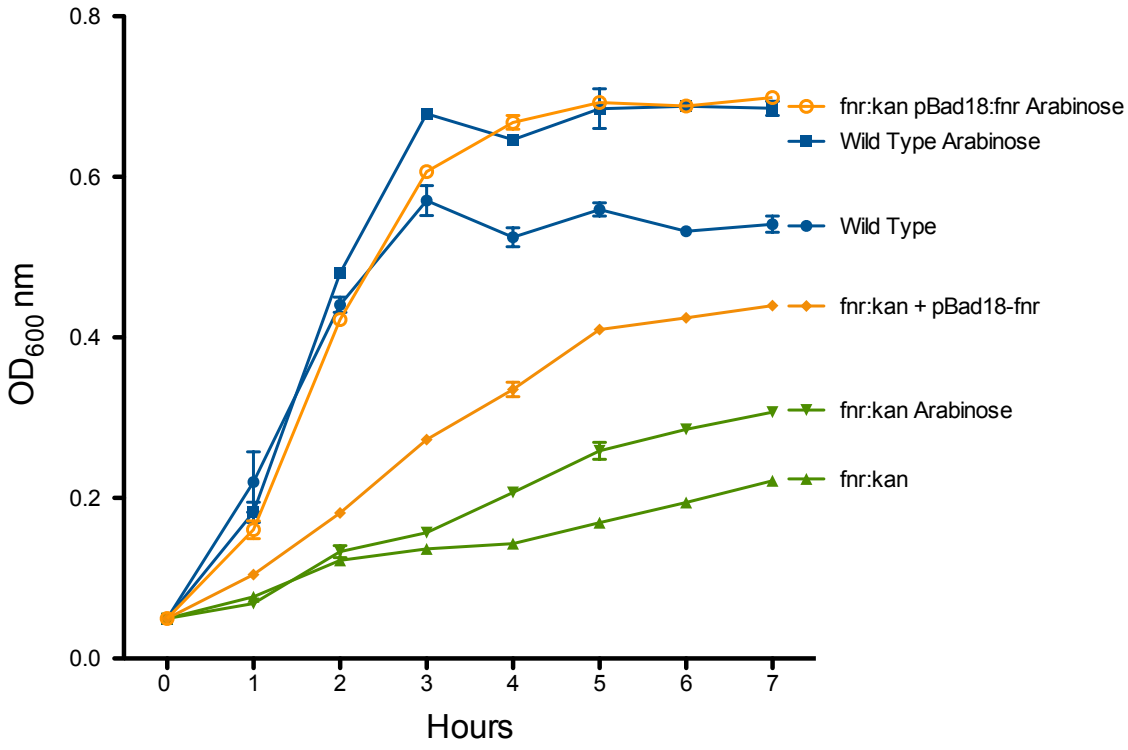



Figure 3.3 Growth of ETEC H10407 in anaerobic conditions

All growths represent 3 biological replicates grown in 5 mL CFA broth at 37°C and error bars represents standard error of the mean. Arabinose (0.2%) was added when indicated to induce expression of *fnr* in the pBad18-*fnr* complementation plasmid.

However, in controlled *in vitro* growths, *sta2* gene expression is influenced by environmental oxygen while *sta1* is not (Figure 3.2. Although highly homologous, the differences in nucleotide sequence between *sta1* and *sta2* ensured their expression could be independently determined when using our mapping parameters, and contrasting regulation between these two genes has been previously established in promoter::lacZ fusion assays[129]. Overall, we found that simply altering environmental oxygen alone can greatly influence virulence factor expression in ETEC, with CFA, heat stable, and heat

labile toxins genes all showing altered expression in response to changes in environmental oxygen (Figure 3.2).

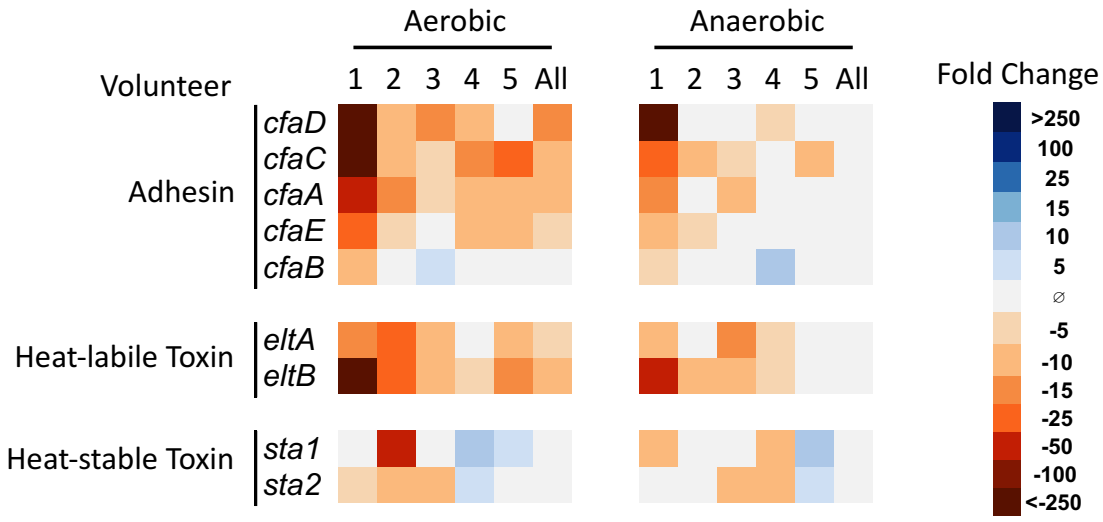



Figure 3.4 Virulence gene expression per volunteer sample.

The transcriptome of each volunteer sample was compared to *in vitro* growth in CFA broth under either aerobic or anaerobic atmospheres (3 biological replicates per *in vitro* condition). Classical virulence factor expression is noted per volunteer sample (numbered 1-5) and together as an *in vivo* group representative of all five samples ("All"). Significant fold change was considered > |3| with a false discovery rate correct p value of <0.05 as outline in the methods.

### 3.2.3 The FNR Transcriptional Regulator Globally Influences Virulence Factor Expression

To better understand how ETEC sensing of environmental oxygen is linked with virulence gene expression we considered the oxygen-sensitive transcription factor FNR (Fumarate Nitrate reduction Regulator). FNR is a well characterized dimeric transcription factor which can be directly inactivated by oxygen[185,186]. In oxygenated conditions, an exposed Fe-S cluster in FNR is oxidized which eliminates the ability of

FNR monomers to dimerize. The activity of antioxidant enzymes such as IscS can reverse this process, restoring FNR protein activity[190]. In this way FNR activity is directly sensitive to environmental oxygen in a reversible manner.

We therefore hypothesized the anaerobic virulence factor repression was dependent on FNR activity. To test this we used recombineering to produce a *Δfnr* H10407 strain and compared the transcriptomes of three biological replicates of WT and *Δfnr* H10407 grown anaerobically to mid-log phase in CFA broth. We found a global increase of ETEC virulence factor expression in the *Δfnr* background (Figure 3.2). Interestingly, both *sta* genes were upregulated in this background, indicating that FNR can influence their expression despite limited changes between aerobic and anaerobic growth in WT bacteria.

### 3.2.4 FNR Binds to Classical Virulence Factor Promoter Regions

To examine if FNR was directly influencing virulence factor gene expression we used the online database Prodoric[191] to determine if there were FNR binding sites in the upstream regions of virulence genes. We found multiple predicted FNR binding sites in the promoter regions of all classical virulence genes (Figures 3.5-3.8). Using electrophoretic mobility shift assays, we found FNR was capable of binding to the promoter regions of *cfaD*, *eltA*, *sta1* and *sta2 in vitro* (Supplementary Figures 3-6). This is consistent with the *Δfnr* RNA-sequencing data and direct FNR repression of classical virulence genes and operons.

However, FNR has a well-defined role in regulating low-oxygen metabolism genes and therefore has a broad influence on cell physiology (Supplemental Figure 7)[185]. Consequently, it is possible that FNR is additionally altering virulence factor expression by influencing other known or unknown regulatory factors, although our data suggest this

is unlikely. Sensing of glucose concentrations via the transcription factor CRP was recently shown to influence transcription of promoter::lacZ fusions, and in Enterohemorrhagic *E. coli*, CRP expression increases when grown anaerobically[192]. However, we did not find differential expression of CRP (*ETEC_3608*) when WT ETEC or *Δfnr* was grown anaerobically.

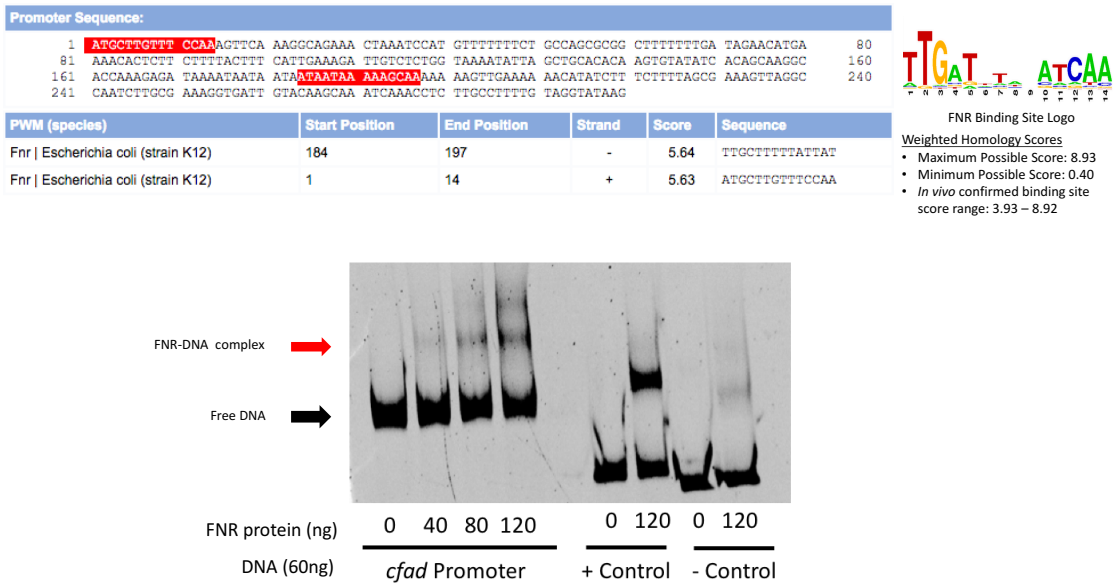



Figure 3.5 *In vitro* binding of FNR to the *cfaD* promoter.

**a**. Predicted FNR binding sites in the promoter region of *cfaD* using the online software Prodoric Virtual Footprint Promoter Analysis Version 3.0 online tool to predict *E. coli* K12 FNR binding sites. Binding sites are highlighted in red, and a score indicating how closely the binding sites match the FNR binding site consensus logo are given. A perfect match to the consensus sequence scores a 8.93, and *in vivo* confirmed binding sites have a score range between 3.93 and 8.92 as determined by Prodoric. **b**. Electrophoretic Mobility Shift Assay (EMSA). Purified oxygen stable FNR variant $(FNRD154A)_2$-His6 was mixed at the indicated amounts with PCR amplified *cfaD* promoter DNA (500 base pairs long starting at the -1 position) to test for FNR's ability to bind to promoter DNA. A 500bp region was used to be sure the promoter for *cfaD* was correctly PCR amplified instead of the similar promoter region for the *cfaD2* homolog found elsewhere. FNR-DNA complexes have hindered movement in a 6% polyacrylamide DNA retardation gel which can be seen as a shift in the mobility of the DNA. $(FNRD154A)_2$-$His_6$ efficiently bound to the *cfaD* promoter and a previously characterized positive control promoter (*ydfZ* gene promoter), but not to a negative control random sequence DNA probe without FNR binding sites.

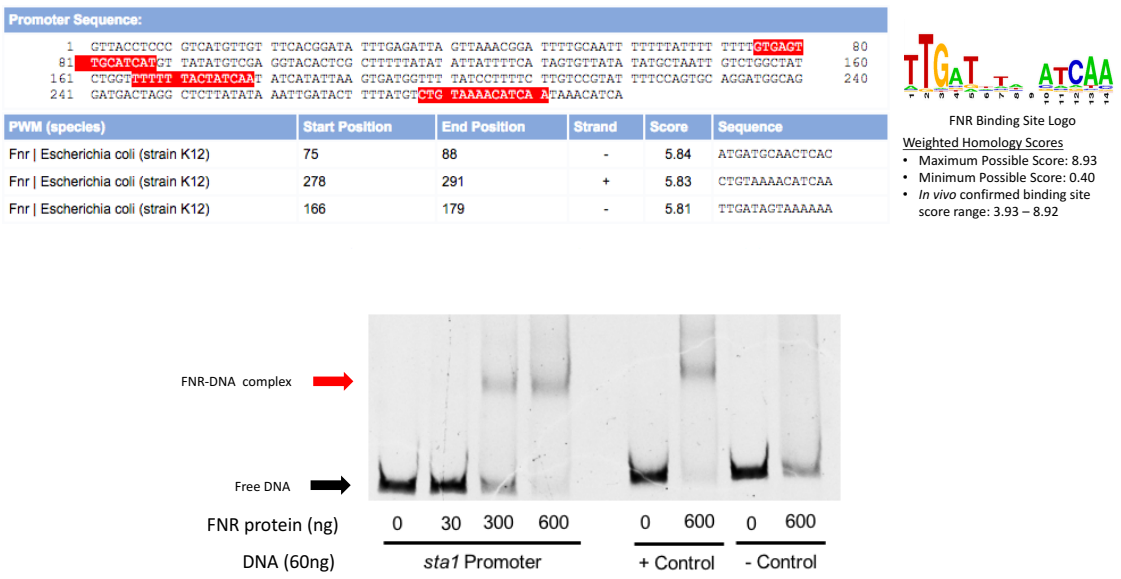



Figure 3.6 *In vitro* binding of FNR to the *sta1* promoter.

**a**. Predicted FNR binding sites in the promoter region of *sta1* using the online software Prodoric Virtual Footprint Promoter Analysis Version 3.0 online tool to predict *E. coli* K12 FNR binding sites. Binding sites are highlighted in red, and a score indicating how closely the binding sites match the FNR binding site consensus logo are given. A perfect match to the consensus sequence scores a 8.93, and *in vivo* confirmed binding sites have a score range between 3.93 and 8.92 as determined by Prodoric. **b**. Electrophoretic Mobility Shift Assay (EMSA). Purified oxygen stable FNR variant $(FNRD154A)_2$-His6 was mixed at the indicated amounts with PCR amplified *sta1* promoter DNA (300 base pairs long starting at the -1 position) to test for FNR's ability to bind to promoter DNA. FNR-DNA complexes have hindered movement in a 6% polyacrylamide DNA retardation gel which can be seen as a shift in the mobility of the DNA. $(FNRD154A)_2$-$His_6$ efficiently bound to the *sta1* promoter and a previously characterized positive control promoter (*ydfZ* gene promoter), but to not the *sta2* coding sequence used as a negative control (300bp).

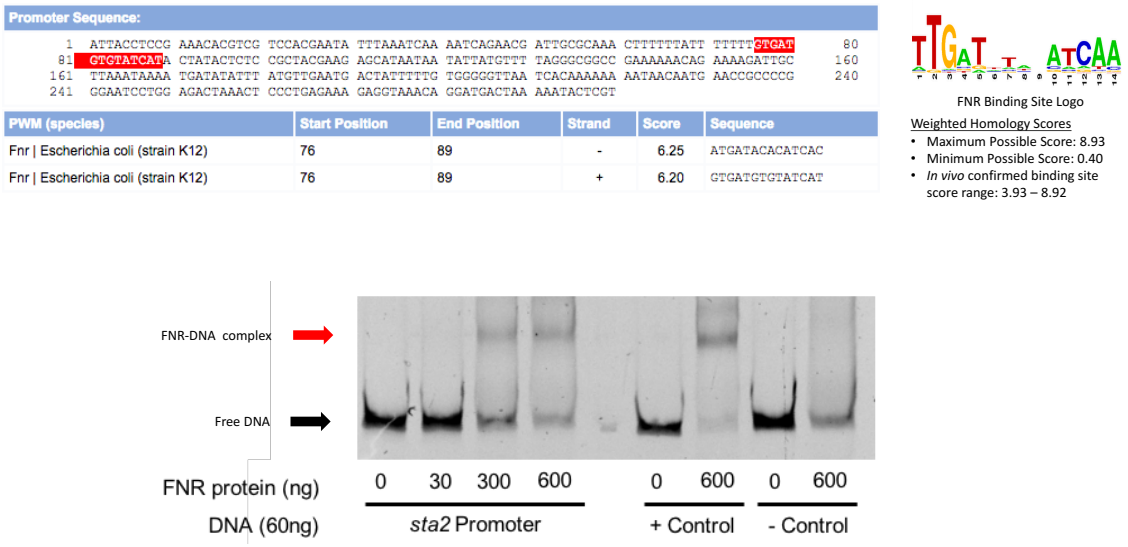



Figure 3.7 *In vitro* binding of FNR to the *sta2* promoter.

**a**. Predicted FNR binding sites in the promoter region of *sta2* using the online software Prodoric Virtual Footprint Promoter Analysis Version 3.0 online tool to predict *E. coli* K12 FNR binding sites. Binding sites are highlighted in red, and a score indicating how closely the binding sites match the FNR binding site consensus logo are given. A perfect match to the consensus sequence scores a 8.93, and *in vivo* confirmed binding sites have a score range between 3.93 and 8.92 as determined by Prodoric. **b**. Electrophoretic Mobility Shift Assay (EMSA). Purified oxygen stable FNR variant $(FNRD154A)_2$-$His_6$ was mixed at the indicated amounts with PCR amplified *sta2* promoter DNA (300 base pairs long starting at the -1 position) to test for FNR's ability to bind to promoter DNA. FNR-DNA complexes have hindered movement in a 6% polyacrylamide DNA retardation gel which can be seen as a shift in the mobility of the DNA. $(FNRD154A)_2$-$His_6$ efficiently bound to the *sta2* promoter and a previously characterized positive control promoter (*ydfZ* gene promoter), but to not the *sta2* coding sequence used as a negative control (300bp).

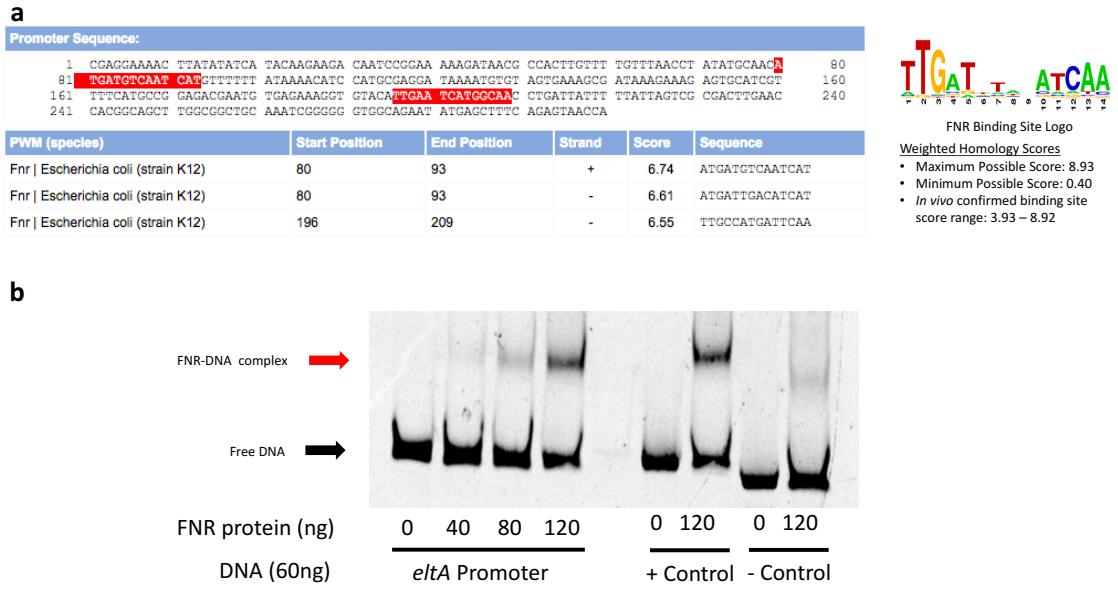



Figure 3.8 *In vitro* binding of FNR to the *eltA* promoter.

*In vitro* binding of FNR to the *eltA* promoter. **a**. Predicted FNR binding sites in the promoter region of *eltA* using the online software Prodoric Virtual Footprint Promoter Analysis Version 3.0 online tool to predict *E. coli* K12 FNR binding sites. Binding sites are highlighted in red, and a score indicating how closely the binding sites match the FNR binding site consensus logo are given. A perfect match to the consensus sequence scores a 8.93, and *in vivo* confirmed binding sites have a score range between 3.93 and 8.92 as determined by Prodoric. **b**. Electrophoretic Mobility Shift Assay. Purified oxygen stable FNR variant $(FNRD154A)_2$-His6 was mixed at the indicated amounts with PCR amplified *eltA* promoter DNA (300 base pairs long starting at the -1 position) to test for FNR's ability to bind to promoter DNA. FNR-DNA complexes have hindered movement in a 6% polyacrylamide DNA retardation gel which can be seen as a shift in the mobility of the DNA. $(FNRD154A)_2$-$His_6$ efficiently bound to the *eltA* promoter and a previously characterized positive control promoter (*ydfZ* gene promoter), but not to a negative control random sequence DNA probe without FNR binding sites.

Osmolarity was also shown to regulate expression of virulence promoter::lacZ fusions via the genome-binding protein HNS[129], and HNS was significantly downregulated 3 fold during anaerobic growth compared to aerobic growth. HNS has been shown to influence FNR and CRP binding to promoter regions[185] and directly represses *eltAB*, *sta1*, and *sta2* promoter::lacZ reporter fusion expression in an *E. coli* K-

12 background[129]. Therefore, decreased expression of HNS during anaerobic growth could relieve this direct repression and result in increased expression of toxin genes. However, only *sta2* showed increased expression during anaerobic growth, and HNS expression was not altered in the *Δfnr* background compared to WT. Interestingly, previous work with promoter::lacZ fusions showed *sta2* was also the only toxin gene shown to be repressed by both direct HNS-dependent repression and HNS-dependent inhibition of promoter activation. This stronger reliance on HNS regulation may explain why *sta2* was the only virulence factor to have increased expression while HNS expression is decreased. Previous work also showed HNS repressed expression of an *eltAB* operon promoter::lacZ fusion in a K-12 background. However, despite a 3 fold decrease in HNS expression during anaerobic growth, we did not see relief of this repression in ETEC, and our data shows the downregulation of *eltAB* is FNR-dependent.

Beyond sensing oxygen, glucose, and osmolarity, ETEC is likely integrating other *in vivo* signals to regulate virulence factor expression, including previously described bile and bicarbonate. Our data shows that without additional manipulations, simply changing environmental oxygen influences adhesin, heat-labile, and heat-stable toxin gene expression, and that deletion of the oxygen-sensitive FNR transcription factor results in a global increase in expression of all classical virulence factor genes.

### 3.2.5 FNR is a Regulator of ETEC Biofilm Formation

Biofilm production provides bacteria protection from environmental stressors inside and outside of the host and has been well described in many pathogens[193]. In ETEC, biofilm production on abiotic surfaces within the homes of afflicted areas is correlated with seasonal ETEC epidemics[84]. While characterizing ETEC in low-oxygen conditions we unexpectedly observed *Δfnr* H10407 spontaneously produced prominent

biofilms during anaerobic overnight growth in CFA broth at 37°C, despite vigorous shaking (Figure 3.9).

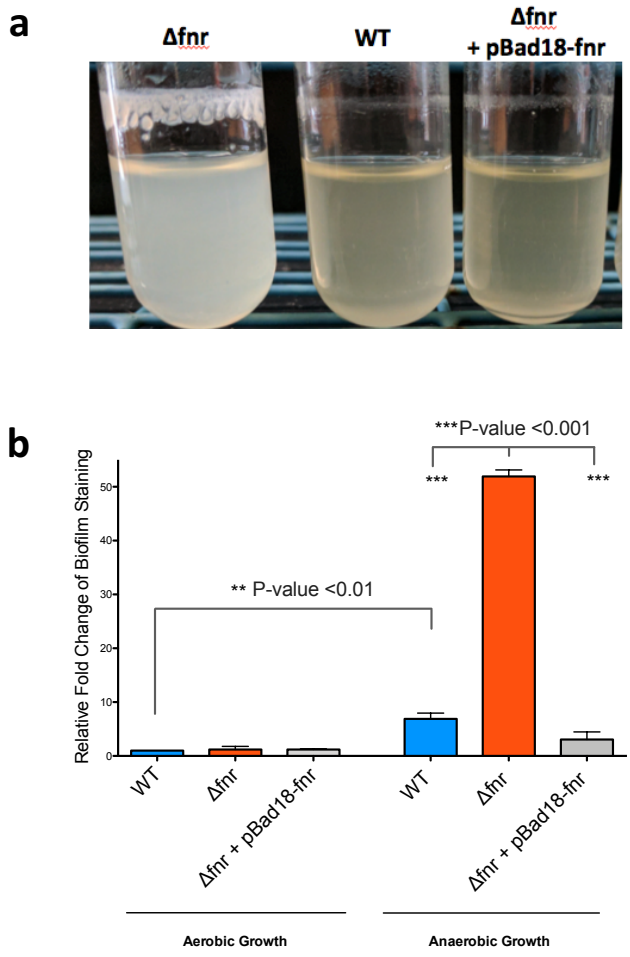



Figure 3.9 FNR regulated biofilm formation in ETEC

**a.** A picture of spontaneously occurring biofilm production when H10407 is grown in anaerobically at 37C with shaking. Image is representative of more than 10 independent replicates. An *fnr* mutant produces more biofilm than wild type and this phenotype can be complemented by expressing the *fnr* gene from the pBad18 plasmid in the *fnr*:kan mutant background. All depicted growths included 0.2 % arabinose for consistency with the induced complemented strain. **b.** Quantification of ETEC biofilm formation exemplified in figure 5a using optical density readings of crystal violet stained biofilms as described in the methods section. All growths represent three independent biological replicates grown with 0.2 % arabinose for consistency with the induced complementation strain. The relative amounts of biofilm produced by various ETEC H10407 strains in the indicated conditions are shown compared to WT ETEC H10407 biofilm production, which is set as 1 as a reference. Statistical significance determined by one-way ANOVA with a Tukey multiple comparison test and error bars represent the standard error of the mean.

Quantitation of biofilm production via crystal violet staining revealed biofilm production increased 10 fold compared to the WT, and this phenotype was complemented

when FNR expression was restored via expression from an inducible plasmid. More broadly, we found WT ETEC H10407 biofilm production increased 5 fold when grown anaerobically compared to aerobic growth.

As expected, the FNR mutant produces WT levels of biofilm when grown aerobically. However, anaerobic biofilm production in the *fnr* mutant was increased 50 fold compared to the WT strain grown aerobically. Together these data depict an interesting dynamic where ETEC can readily grow as a biofilm in an anaerobic environment, yet activation of FNR in this condition limits the extent of biofilm production.

Bacteria rely on a variety of mechanisms to form biofilms in response to changes in their environment[194]. To identify FNR influenced genes that contribute to anaerobic biofilm formation in ETEC we sought to identify upregulated genes in *Δfnr* H10407 that have been previously associated with biofilm formation. The ETEC TibAB adhesin has been shown to contribute to cohesion and biofilm formation[181], but unexpectedly, these genes are resoundingly downregulated in the *Δfnr* strain (Figure 3.2), suggesting they are unlikely to be responsible for increased biofilm formation. Alternatively, the main structural subunit of the Fim adhesin, *fimA*, was upregulated > 10 fold. Fim-based adhesion in ETEC occurs *in vivo* and *in vitro* via a specific binding of the adhesive tip protein FimH to D-mannose[195,196]. Therefore, exogenous D-mannose quenches Fim adhesive ability and disrupts biofilm formation, and sensitivity to D-mannose is frequently used as the standard to examine Fim-dependent biofilm formation or cellular aglutination[196–198]. To see if D-mannose dependent adhesion was contributing to biofilm formation in *Δfnr* H10407 we measured biofilm production in the presence of increasing concentrations of D-mannose. A concentration dependent significant decrease

in biofilm production was seen in the presence of D-mannose, indicating the Fim adhesin contributes to biofilm formation in ETEC (Figure 3.10).

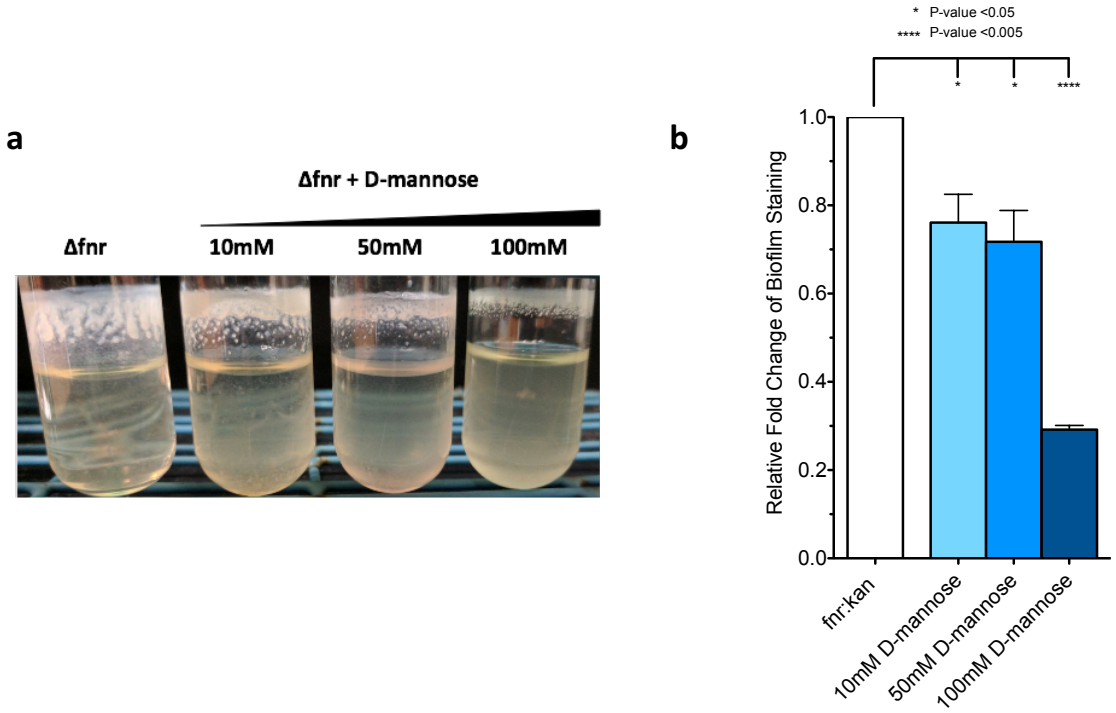



Figure 3.10 Exogenous D-mannose influence on ETEC biofilm formation

**a**. Picture of ETEC biofilm formation of an *fnr*:kan strain grown with increasing amounts of D-mannose to quench activity of the fimbrial adhesin protein *fimH*. FimH has been shown previously to contribute to adhesion in ETEC. Representative of at least 3 independent growths. **b**. Quantification of ETEC biofilm formation exemplified in figure Supplementary Figure 5a using optical density readings of crystal violet stained biofilms as described in the methods section. All growths represent three independent biological replicates. Statistical significance was determined using a one-way ANOVA test with tukey multiple comparisons test. Error bars represent standard error of the mean.

### 3.2.6 The ETEC Transcriptome in Human Infection Samples

Beyond virulence factors, we defined ETEC genes and pathways that were differentially regulated *in vivo* based on comparisons to different environmental oxygen conditions. Over one thousand genes are differentially regulated *in vivo* compared to the

atmospheric conditions we tested (Figure 3.1a). *In vitro* microaerobic growth most closely matched the infected stool transcriptome, while aerobic growth was the most different. (Figure 3.1a). A statistical analysis of pathways that are differentially regulated *in vivo*, per environmental oxygen comparison, show virulence factor expression differences were highly dependent on atmospheric oxygen growth while chemotaxis and motility pathways were not (Figure 3.1b).

Despite the differences between atmospheres, 361 genes were differentially regulated *in vivo* independently of environmental oxygen changes (Figure 3.1a). Shared functional roles clearly emerged while examining these data by hand (Figure 4).

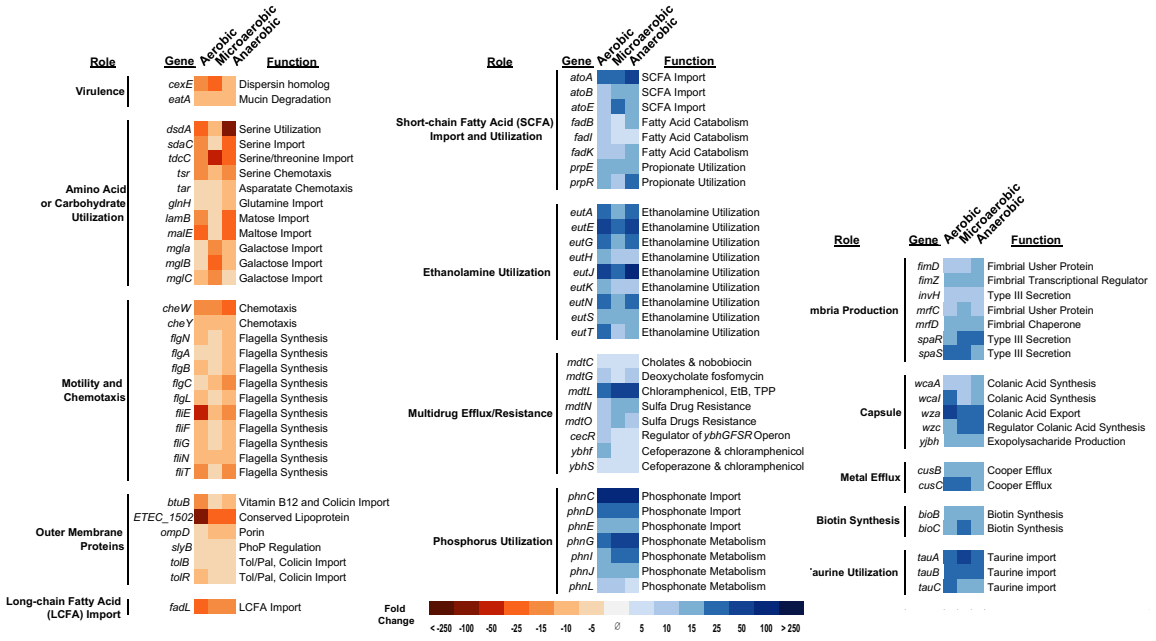



Figure 3.11 In vivo differentially expressed genes independent of atmosphere

ETEC genes that showed differential gene expression *in vivo* despite the *in vitro* atmosphere comparisons (fold change > |3|, false discovery rate corrected $p < 0.05$) manually curated to highlight genes with clear functional roles. All changes in gene expression can be found in Supplemental Table 1.

We found many nutrient acquisition pathways that have been important for intestinal colonization in small animal models are significantly differentially regulated in human infection samples. For example, short chain fatty acid utilization has shown to be crucial for colonization and/or survival of some intestinal microbes, and we see short chain fatty acid utilization genes in ETEC are strongly upregulated in human infection[199]. Congruently, the long-chain fatty acid import protein *fadL* is downregulated *in vivo*.

Similarly, striking regulation is seen in ethanolamine utilization, with upregulation of 9 genes involved in uptake and utilization of ethanolamine during human infection. Host intestinal epithelial cell turnover releases ethanolamine into the intestinal lumen, and in Enterohemorrhagic *E. coli* and Salmonella, ethanolamine is used as a nitrogen source to create a competitive advantage over the microbiota to establish an infection[200,201]. These data jointly suggest ETEC may be utilizing ethanolamine during human infection in a similar manner. Biotin synthesis was also upregulated *in vivo*, and in Salmonella, biotin synthesis is upregulated upon epithelial cell infection and deletion of the biotin synthesis gene *bioB* resulted in attenuated mouse colonization[202].

Beyond nutrient acquisition genes, many genes that play a role in resistance to antibiotics or host factors are differentially regulated *in vivo* (Figure 3.11). These include the mucin degradation gene *eatA*; fimbrial genes *fimD*, *fimZ*, *mrfC* and *mrfD*; and invasion/secretion associated genes *invH*, *mrfC*, and *mrfD,* and *cexE*. Genes involved in multidrug efflux and resistance were also upregulated, including bile acid and nobobiocin/fosfomycin resistance genes *mdtC*/*mdtG*, Sulfa drug resistance genes *mdtN* and *mdtO*, and cefoperazone and chloramphenicol resistance genes *ybhF* and *ybhS*.

### 3.2.7 Common Gene Regulation Between UPEC and ETEC During Human Infection

Few transcriptomes of bacterial pathogens in human infection samples exist. An exemplary data set is that of uropathogenic *E. coli* (UPEC)[135], where an infected urine sample typically represents a monoculture of urinary tract infecting bacteria. This minimally complex sample enables efficient sequencing and analysis of *in vivo* pathogen transcriptomes. Interestingly, many of the upregulated ETEC genes in infected stool we also upregulated in infected urine samples collected from human UPEC infections[135]. Notably, ethanolamine utilization, phosphonate utilization, taurine import, colonic acid synthesis, and copper efflux were all major upregulated pathways despite the differences between these pathogens and site of infection. As most *E. coli* urinary tract infections are established by *E. coli* found in fecal flora, these data suggest a common transcriptional response to the host across *E. coli* strains and infection sites in humans.

## 3.3 DISCUSSION

We examined ETEC H10407 directly in human infection samples from a controlled human infection model and found atmospheric oxygen influences virulence gene expression via the FNR transcription factor. Previously, ETEC has not been well characterized in low-oxygen conditions and the role of FNR in ETEC virulence and physiology was unknown.

We found CFA/I adhesin genes, including the master regulator *cfaD*, and the heat-labile toxin genes *eltAB,* are downregulated when ETEC is grown anaerobically. Consistent with FNR-dependent repression, these genes are upregulated in a *Δfnr* strain, including both heat-stable genes in H10407, *sta1* and *sta2*. This strikingly shows that a single regulator in ETEC can globally influence virulence factor expression, and *in vitro*, FNR can bind to the promoter regions of each of these genes. Similarly, anaerobic growth

*in vitro*, which represents the near zero oxygen availability status of distal small intestine and anaerobic large intestine, as well as bacterial overgrowth in the proximal small intestine, mimicked the virulence gene expression profile seen in stool samples.

Low oxygen availability is a hallmark of the intestinal environment, and *in vivo*, *E. coli* can respond to their proximity to the epithelium by sensing oxygen seepage from host cells[188]. Together these data suggest ETEC virulence factor expression may be repressed by FNR in the low-oxygen lumen, and that repression is relieved when ETEC nears the intestinal epithelium and oxygen inactivates FNR (Illustration 3.1).

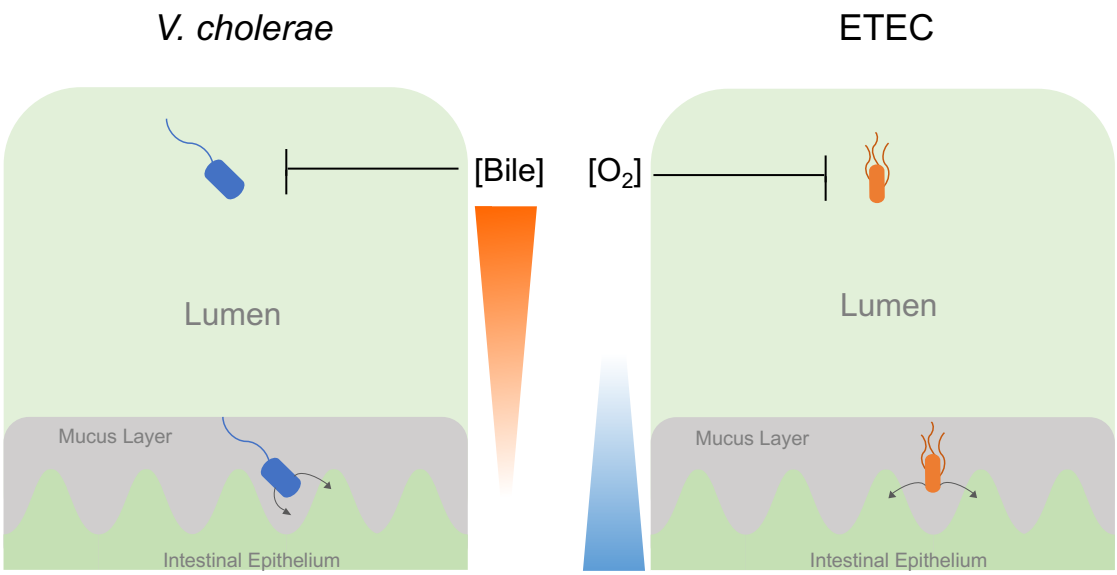



Illustration 3.1 A model of oxygen-dependent ETEC virulence gene regulation.

Our study suggests ETEC directly senses environmental oxygen via the oxygen-sensitive transcription factor FNR which results in repressed virulence gene expression in low oxygen environments like the intestinal lumen. When ETEC nears the intestinal epithelium, where oxygen seepage from host cells creates a zone of increased oxygen concentrations, FNR is inactivated which relieves virulence gene repression nearest to the host epithelium. Other intestinal pathogens, such as *Vibrio cholerae*, use a similar proximity-sensitive approach to regulating virulence genes within the host. In *V. cholerae*, bile has been show to repress cholera toxin expression, and similar to oxygen, bile exists as a gradient with the intestinal environment. In both cases, concentrations of host derived factors result in repression of virulence gene expression in the lumen and relief of the repression when in close proximity to the target epithelium.

Convergent evolution across intestinal pathogens appears to have favored similar methods of host-cell proximity sensing. For example, the diarrheal pathogen *Vibrio cholerae*, also senses a host cue (bile) to repress cholera toxin expression in the lumen[203], which is then relieved when in close proximity to host cells (Figure 5).

FNR and the ability to switch between aerobic and anaerobic metabolism is known to be essential for *E. coli* colonization[204], and other intestinal pathogens also utilize FNR to regulate select colonization factors[188]. The FNR regulon has been well defined in laboratory *E. coli* K-12 and interplays with other regulator proteins such as HNS and the related CRP transcriptional regulator which responds to cyclic AMP in low-glucose conditions[185]. Recent work has demonstrated the activity of both HNS and CRP contribute to virulence gene expression in aerobic conditions using promoter::lacZ fusion assays in K-12 background[129], and certainly, all may play a role in the low-glucose, low-oxygen environment of the intestinal tract.

Beyond classical virulence factor expression changes, we found many cellular processes were differentially regulated *in vivo*. These include ethanolamine utilization, short chain fatty acid utilization, phosphonate utilization, and copper efflux. Notably, previous work demonstrated these same processes were also upregulated in human UPEC infections. Further, all of these have been shown to contribute to fitness in small animal models of infection, and here we found they respond transcriptionally to the human host during ETEC infection.

Besides offering insight into ETEC gene regulation *in vivo*, these data have implications for studies that examine ETEC virulence factor expression in collected infected samples[205,206]. Our data suggest stool samples should be maintained or preserved with consideration for atmospheric oxygen to be sure differences in virulence factor expression between samples or strains cannot be attributed to differences in oxygen exposure after collection.

During our characterization of ETEC in low-oxygen conditions we noticed an oxygen-sensitive biofilm production phenotype. Only a few studies have investigated ETEC biofilm formation, although ETEC biofilms in home water reservoirs in afflicted

areas are correlated with seasonal epidemics[84]. Here we show robust ETEC biofilms can be produced overnight when *Δfnr* H10407 is grown anaerobically in CFA media. Together, this demonstrates an interesting FNR-dependent link between ETEC's lifestyle in the environment and in the host. This study represents the joint efforts between clinicians and researchers to define an *in vivo* perspective of a worldwide human pathogen.

# Chapter 4: Discussions on Bacterial Adaptation During Human Infections

## 4.1 Insights into *C. jejuni* Infection in Humans and Future Work

The work described in Chapter 2 has significantly advanced our understanding of *C. jejuni* infections in humans. First, *C. jejuni* infections are associated with an assortment of possible symptoms, each with seemingly unpredictable severity between patients. Previously reported controlled human infection models revealed that even when a group of volunteers are infected from the same inoculum at the same time the severity of symptomatic disease can markedly vary between volunteers.

Considering *C. jejuni*'s phase variable nature, one hypothesis to explain disease severity discrepancies between humans is that more virulent genetic backgrounds take hold in particular patients by chance. This could be due to either 1) genetic drift of a more virulent phenotype 2) chance bottleneck colonization[52] by bacteria with a differing virulence between volunteers or 3) chance fluctuations in phase variant frequencies of virulence genes *in vivo* that result in differing symptomatic disease. This hypothesis therefore focuses on bacterial contributions to disease severity in humans. An alternative hypothesis is that host factors dictate symptom severity. In this case, since all volunteers are infected from the same inoculum, the symptom variability between patients could result from 1) differing immune responses per host i.e. a stronger intestinal inflammatory response may be more likely to produce bloody diarrhea or 2) differing commensal gut flora that influence *C. jejuni*'s ability to induce symptoms or 3) yet to be identified non-immune factors.

Our data clearly demonstrate there is no detectable colonization bottleneck in humans, as a very consistent genetic selection pressure was identified even on the day of inoculation. Had bottlenecking been a major issue, seemingly random genetic variations

would have been seen colonizing different individuals, as only some genetic backgrounds would be able to pass the colonization bottleneck by chance. Further, because very similar genetic variations were identified across patients and throughout the entirety of acute infection, there is little evidence to suggest genetic drift or more virulent phenotypes could produce more severe symptoms. Remarkably, there was no particular genetic variation (phase variation or otherwise) that was statistically significantly associated with samples from volunteers who experienced severe disease. This strongly suggests host factors, and not differences between infection populations, dictate disease severity. Future work should examine if differences between the commensal microbiome populations are associated with volunteers with severe disease.

It is also possible that the bacteria that are shed from the host are not representative of the genotypes that actively induce symptomatic disease. For instance, if invasive bacteria are solely responsible for symptoms, they may not be shed from the host, and therefore we are not identifying them be screening post-infection isolates. This possibility is unlikely during acute infection, as the current literature suggestions invasion is a low-frequency event, so genotypes that promote invasion should still be detected in shed bacteria.

Another potential limitation to this study was the requirement to use pooled infection isolate populations. Infected feces were streaked onto solid medium and single colony isolates were collected and pooled together (per patient sample) for genomic sequencing. This enabled efficient sample processing and analysis. However, it makes it difficult to determine the exact genetic variations that existed in a single genome. Knowing which genetic variants were identified within a single genome would elucidate potentially synergistic variants.

Although no genotypes are associated with symptom severity, variability in 11 genetic loci were associated with persistent infections in humans. This is the first description of factors that contribute to recrudescent infections in humans, and unlike symptom severity, there are clear bacteria-specific factors associated with recrudescence. Importantly, humans are the only model system we have to accurately represent recrudescent infections, as small animal models are routinely cured with antibiotics while humans can relapse despite what initially appears to be effective antibiotic treatment. It is likely not a coincidence that phase activation of the *cipA* gene, a known *in vitro* invasin, is associated with persistent bacteria. It strongly suggests that persistent *C. jejuni* are invasive, and it is attractive to speculate that enhanced invasive ability promotes *C. jejuni* to better evade the host immune response and antibiotic exposure. There is also strong selection of flagellar glycosylation genes in persistent bacteria. Although O-linked glycosylation genes have been characterized in *C. jejuni*, their functional role *in vivo* is unclear. Likely scenarios include changing glycosylation patterns helps avoid immune detection of the flagella, and since flagella are associated with motility, attachment, and invasion, phase variation of O-linked glycosylation genes may also contribute to invasion.

Finally, the most surprising genetic association with *C. jejuni* recrudescence is truncation of the CmeR bile-sensing transcription factor. Bile-sensing is a hallmark of intestinal pathogens, and losing the ability to sense changes in environmental bile concentrations is a surprising phenotype for an intestinal pathogen. The truncations identified eliminated the dimerization domain that is essential for CmeR's ability to bind DNA and repress its target promoters. Therefore, truncation would lead to constitutively expression of the CmeR regulon, which includes a bile-acid efflux pump. Future work should determine the extract to which this change in regulon expression influences *C.*

*jejuni*'s ability to survive bile acid stress. It is worth noting that the CmeR genetic variations we identified were in two different non-phase variable loci and occurred in 1/3 of relapse infections tested. These non-phase variable mutations naturally occur at a very low frequency, but given the high frequency they existed in their respective relapse infection populations, they likely conferred a significant fitness benefit to persistence *in vivo*.

We also utilized samples from a non-human primate infection model to further verify genetic selection pressures *C. jejuni* experiences in primates. Strain CG8421 is not adapted to the new world monkey *A. nancymaae* and, therefore, requires a higher inoculation dose to induce symptomatic disease. However, future work should examine if post-infection isolates from this this experiment require a lower infectious dose to induce symptoms. This would suggest the genetic variations we identified do indeed increase success in the host. Further, there are currently no non-human infection models to study recrudescent infections. Future work could examine if infecting *A. nancymaae* with an inoculum enriched for *cipA* "on" bacteria results in recrudescent infection. This would 1) provide greater support for *cipA*'s role in persistence and 2) establish an additional model for studying recrudescent *C. jejuni* infections. Finally, the precise structure of the flagellar modification produced by CipA should be determined and evaluated for its potential to be used as an antigen in future vaccine preparations. Perhaps immunizing against CipA modification may lower the incidence of persistent C. jejuni human infections.

### 4.2 A SIMPLER MODEL OF ETEC VIRULENCE GENE REGULATION

The work presented in Chapter 3 presents a clear, unified regulation scheme for ETEC virulence gene regulation. For the first time, a single environmental factor, and

corresponding transcriptional regulator, was show to influence every classical virulence gene in ETEC. This regulatory factor, oxygen, further provides a simplistic view virulence gene expression: as ETEC approaches the oxygen rich zone nearest to the host epithelium, virulence factor expression is increased.

However, the amount of genetic diversity in ETEC virulence factors is staggering, and it is reasonable that not every ETEC isolate will behave like the prototypical H10407 strain used in this work. For instance, oxygen regulation of ST is irrelevant in an ETEC strain that does not harbor ST, and without ST, there may be less pressure to coordinate remaining virulence factors via a single environmental signal. Promoter analysis in diverse ETEC strains could be used to identify putative FNR binding sites to provide more insight, and EMSA assays were straight forward in determining FNR's ability to bind to such promoters.

FNR is a well-known virulence regulator in other enteric pathogens. Therefore, discovering its role in ETEC virulence gene regulation is not a controversial mechanism that requires unusually burdensome experimentation. However, future work examining the role of FNR in additional ETEC strains will be valuable and appreciated. Particularly, the use of chromatin immunoprecipitation of FNR to identify its binding sites *in vivo* (ChIP-seq) is a clear next step. ChIP-seq has been used to characterize the FNR regulon in laboratory K-12 *E. coli* and this data set can be used as a positive control to verify an ETEC ChIP-seq experiment identified known chromosomal FNR binding sites in E. coli – for instance, in the promoter region of the *ydfZ* gene which contains a conserved FNR binding site sequence and used as a positive control in the EMSA assays prestend in Chapter 3. ChIP-seq data would also be beneficial to confirm that FNR is displacing HNS and/or CRP binding during anaerobic growth.

Finally, the new FNR and oxygen-dependent regulation scheme paired with within-human expression data can reveal genes that are co-expressed with ETEC virulence factors. For instance, identifying surface exposed gene products (which would be accessible to the immune system or small molecule therapies) that are 1) highly expressed in concert with virulence genes *in vivo* and 2) are well conserved across ETEC strains, may identify new proteins that could potentially be used in future ETEC vaccines. It is important to note that major virulence factors, such as adhesins, are highly variable across ETEC strains, and therefore more highly conserved targets (which still share expression patterns with known virulence factors) could prove useful in making a broadly protecting ETEC vaccine.

### 4.3 Creating a Framework for Future Clinical Trials

The work presented in Chapters 2 and 3 relied on preserved human infection samples taken from controlled human challenge models. These challenges are used to establish minimum infectious dosages for pathogens, characterize host symptoms and immune responses, and then later to test the efficacies of new potential therapies. Typically, the focus of these studies are clinician-driven care for volunteers and immunological studies on the host response to the infection and the treatment. Here we have expanded the scope of these trials to include comprehensive analysis of the pathogen itself. The most striking benefit to studying pathogens in controlled human infections versus community infections is the inoculum consistency. Every volunteer ingests a portion of a single inoculum, and therefore confounding factors due to wildtype strain variation, differences in inoculum dosages, and differences in time-from-exposure between humans can be eliminated.

As detailed in Chapter 5, this required efforts to training personnel to collect and preserve both inoculum and infection isolate populations and additional microbiologists to analyze these samples. However, given the data generated from this work and presented in Chapters 2 and 3, the effort provided invaluable insight into human infection that would have otherwise gone unrealized. Future clinical trials should consider using a similar sample collection and analysis framework to gain similar useful insights.

Importantly, future trials should integrate the immunological and clinical prognoses observations with observed changes in the pathogen. For instance, before infection, feces samples could be collected to detect changes in the microbiome before, during, and after infection. Similarly, if a volunteer required intervention, such as intravenous rehydration therapy, special attention could be given to examining how the pathogen adapted *in vivo* before and after the therapy. Although this approach could be perused after a trial has completed by synthesizing available data, the most appropriate samples may not have been collected during the trial. For example, if only one sample per volunteer is preserved per day, a sample collected in the morning may not be representative of the disease state 18 hour later that required clinical intervention. Therefore, frequent and dynamic sample collection, conducted while considering the disease state of each volunteer, could be a valuable change in future work.

# Chapter 5: Experimental Methods

## 5.1 Experimental Methods Used in Chapter 2

### 5.1.1 *C. jejuni* Infected Human Feces Sample Preparation for RNA-Sequencing

*Campylobacter jejuni* strain CG8421 infected diarrhea was weighed and added 1:1 in RNA Later reagent as quickly as possible after it was produced and frozen at -80°C. Samples came from volunteers who were not receiving antibiotics and therefore represent untreated infections. Preserved samples were thawed on ice and total RNA extracted via Trizol-chloroform phase separation. DNA was removed (Turbo DNA-free DNAse kit, Ambion), rRNA depleted (Ribo-Zero human and bacterial rRNA removal kits, Illumina), and libraries built for Illumina Sequencing (Ultra Directional RNA Library Prep Kit for Illumina, New England Biolabs).

### 5.1.2 *In vitro C. jejuni* Sample Preparation for RNA Sequencing

*Campylobacter jejuni* strain CG8421 was grown at 37°C in a microaerobic atmosphere (85% nitrogren, 10% carbon dioxide, 5% oxygen) in a Coy Laboratories atmosphere controlled chamber. For RNA-seq of samples grown on agar plates: Mueller-Hinton agar or 5 percent sheep blood agar plates were streaked in triplicate from a frozen stock of *C. jejuni* strain CG8421. >10 Colonies from these plates were re-streaked on fresh plates and mid-log phase bacteria were harvested and suspended in Trizol reagent for RNA extraction. For RNA-seq of samples grown in liquid broth: 5 mL aliquots of Mueller-Hinton broth were equilibrated in the microaerobic chamber for 24 hours before being inoculated and were grown with shaking. First, Mueller-Hinton agar plates were steaked in triplicate from a frozen stock of *C. jejuni* strain CG8421. Colonies from these

plates were used to inoculate equilibrated broth and grown with shaking to stationary phase. These cultures were then used to inoculate freshly equilibrated Mueller-Hinton broth aliquots, which were grown with shaking until mid-log phase. Bacteria were then harvested via centrifugation and resuspended in Trizol reagent for immediate RNA extraction. Extracted RNA was DNAse treated (Turbo DNA Free DNAse Kit, Ambion), rRNA was removed (Ribo-Zero Bacterial rRNA Removal Kit, Illumina) and the samples were prepared for Illumina Hiseq using the TruSeq Stranded mRNA Library Kit (Illumina).

### 5.1.3 Illumina RNA Sequencing and Analysis

For RNA sequencing of samples from infected human feces, three RNA libraries built from the infected feces samples of different volunteers were sequenced using Illumina Hiseq Single-End 50bp reads. Approximately 13 billion reads were obtained in total across all three samples. Using CLC Genomics Workbench software, reads were trimmed to 21bp and aligned to the CG8421 published reference genome using 100 percent exact homology match mapping parameters. At least 3 million reads mapped to the reference genome per sample. For *in vitro* samples, RNA libraries were sequenced and aligned in an identical manner, with at least 15 million reads mapping to the reference genome per sample. To determine differential expression between samples, Reads Per Kilobase per Million reads (RPKM) values were determined for each gene, normalized by quintile, and then were used in a Baggerley's test to determine the weighted proportions fold change between groups of biological replicates (all groups were in triplicate) with a false discovery rate corrected (FDR) p-value. Fold changes $\geq |3|$ with a FDR p-value of $\leq 0.05$ were considered significant. Annotation of differentially

regulated CG8421 genes was made by considering the annotations of homologous genes in strains 81-176 and 11168 as determined by NCBI's online blastp suite. Additional functional annotations were made after considering published literature, using NCBI's PubMed literature search online. Comparison of the CG8421 transcriptome in human feces to the *C. jejuni* 81-176 transcriptome in the chicken cecum was made by using the published list of expression fold changes per 81-176 gene in the chicken cecum as described in Sup. Table 2. Both datasets use mid-log phase microaerobic growth in Mueller-Hinton broth as the reference transcriptome. We compared the list of differentially regulated CG8421 genes in human feces (>|3| fold, <0.05 FDR p-value) to the list of differentially regulated 81-176 genes (>|3| fold, < 0.05 adjusted p-value) in the chicken cecum. Only genes with a homolog in both strains (as determined by similarity of predicted amino acid sequence) were included in the comparison.

### 5.1.4 Human Infection and Inoculum Preservation

Samples were obtained from a previously described[61] controlled human infection model (ClinicalTrials.gov Identifier NCT02280044, Navy Medical Research Center IRB NMRC.2014.0013, and Western Institutional Review Board study number WIRB 1147996). Volunteers were healthy adults recruited from the mid-Atlantic area, signed informed consent prior to participating in the study, and allowed for all future use of the samples obtained during the study. Volunteers who were admitted to the inpatient facility at Johns Hopkins University Center for Immunization Research for the study had ages ranging from 18-50 years old (median 30), normal stool patterns, were not using antidiarrheal or antacid therapy, and had no history of clinically significant diseases. Subjects with a history of *Campylobacter* exposure, a personal and/or family history of

Guillain-Barre syndrome and/or inflammatory arthritis, were positive for HLA-B27, or whose serum immunoglobulin A titer to CG8421 glycine extract was greater than 1:4000 were excluded from the study. Admitted volunteers were inoculated on Day 0 with a *C. jejuni* CG8421 liquid inoculum that was generated using Good Manufacturing Processing (GMP) and kept on ice. The inoculum is derived from Mueller-Hinton agar plate growths that are suspended in phosphate buffered saline, and therefore has complex standing genetic variation (detailed in Sup. Table 3) and is not derived from a single colony. This resuspension method was used for the *Aotus nancumaae* inoculum preparation as well. Immediately prior to inoculation, the PBS resuspension human inoculum (target concentration of $5x10^5$ colony forming units per ml (CFU/mL) was diluted in sodium bicarbonate buffer ($NaHCO_3$) to neutralize stomach acids and ingested. After the last patient was inoculated, a portion of the remaining inoculum was used to verify the infection dosage via growth and colony enumeration on Campylobacter CVA (cefoperazone, vancomycin, and amphotericin B) agar (dosage was $1.7x10^5$ CFU) and another portion was frozen at -80°C to later be used for genomic sequencing.

### 5.1.5 Human Infection Isolate Collection/Preservation

*Campylobacter* isolation from infected feces was performed as previously described[61]. Briefly, infected feces were diluted and temporarily incubated in thioglycollate broth, plated on selective *Campylobacter* media, and incubated in a microaerobic atmosphere at 37C. Colonies on these primary isolation plates were screened by eye and all putative *C. jejuni* colonies were collected and pooled together in glycerol media per sample and frozen at -20°C. An effort was made to collect at least 10 colonies per sample, and samples with insufficient amounts of *C. jejuni* isolates were

identified during analysis of the sample genomes (samples with less than 50x coverage of the genome) and removed from the study.

### 5.1.6 Genomic Sequencing

The frozen inoculum and isolate samples were thawed on ice, genomic DNA was extracted (Easy DNA Kit, Invitrogen) from each sample, followed by processing for Illumina NextSeq Single-Read sequencing. All genome analysis was preformed using CLC Genomics Workbench. Reads were aligned to the CG8421 reference genome for each isolate population. Only isolate population samples with at least 50x coverage of the CG8421 genome were included for genomic variant analysis. These samples averaged >1000x genome coverage (Fig. 2.3).

### 5.1.7 Genomic Variant Calling

Mapped reads were locally realigned and genomic variations compared to the reference genome were identified using a no ploidy assumption low frequency variant detection method. For the highest quality data we manually removed clear false positive variant calls, such as SNPs called in extensive pileups of single reads or in regions of unreliable coverage. By calling variants that occurred in at least 1 percent of any sample population, we identified over 600 genomic variants across all samples. This list of variants, including their frequency in the population, per sample, is provided in Sup. Table 1. To enrich this list for meaningful variants that may impact the human disease state, we aimed to identify genome variants that had major changes in frequency between the inoculum and isolate populations, such as a minority variant in the inoculum

becoming a majority variant in an infection isolate population. By defining a variant as a mutation that occurred in at least 25 percent of a sample population, and identifying variants that differed between the inoculum and an isolate population, we achieved this goal (Fig 1.3). For example, using this approach, variants that were only called in an isolate population(s) on average had a 12-fold increase in frequency between the inoculum and the infection isolate populations (Fig. 2.4). These variants represented a minority of the inoculum population (~5 percent on average), but a majority in the isolate population (~60 percent on average across samples) (Fig. 2.4). These variants were used in the analyses throughout the study.

### 5.1.8 Genomic Variant Analysis

Our variant calling procedure identified 48 genomic variants with major changes in frequency between the inoculum and least one isolate population. Based on variant location in the published annotated CG8421 genome, these variants likely affect ~30 annotated genes. In many cases, different variants across samples had an equivalent predicted affect on the same gene (such as an insertion or deletion induced frameshift at the same location in the same gene) and this is shown for each variant in Sup Table 1d. To incorporate these shared variant affects into our analysis, we manually refined the variant lists to reflect the genes affected by each variant (Illustration 2.3, Step 5), so that different variants with the same affects were considered as equivalent. This enabled us to determine if variation in a particular gene was associated with the human disease state (Illustration 2.3, Step 6) such as if a variation in a particular gene is associated with relapse isolate populations compared to primary infection isolate populations. To do this, we preformed Fisher's exact test with a false discovery rate (FDR) corrected p-value in

CLC Genomics Workbench to compare variants between sample groups. The Microbial Genomics module of CLC Genomics Workbench was used to generate the SNP-tree seen in Figure 4a using the work done by Kass *et al*[207]. Briefly, the tree visually represents how similar genomic variations are between samples. The closer two samples are to each other, the more similar the genomic variations. When examining genomic variants between samples, different variant alleles at the same genomic location are scored as different variants. Other statistical tests used to compare variant frequencies were performed in the statistics and graphing software Prism 6 and are noted in the appropriate figure legends.

### 5.1.9 *Aotus nancymaae* Infection, Isolate Collection, and Analysis

The *Aotus nancymaae* infection protocol was approved by the Naval Medical Research Unit-6 Institutional Animal Care and Use Committee (IACUC approval number NAMRU6-16-03) in compliance with all applicable Federal regulations governing the protection of animals in research. *Aotus nancymaae* monkeys were infected as described previously[62,208] with a single *C. jejuni* CG8421 liquid inoculum that was harvested from Mueller-Hinton agar plates and resuspended in PBS similarly to harvesting bacteria for the human inoculum. Animals were treated with rantidine and CeraVacx to neutralize stomach acid and were anesthetized with ketamine intramuscularly. Animals were inoculated with 5 x $10^{11}$ bacteria in 5 ml of PBS orogastrically for infection and a portion of the inoculum was frozen at -80°C for preservation and genomic sequencing. *C. jejuni* infected diarrhea was plated on *Campylobacter* selective media as previously described[62,208] and isolates were collected and pooled per sample, processed for sequencing, and data analyzed in a parallel manner as described for the human infection

isolate population samples. After genomic sequencing, 12 diarrheal samples from 5 monkeys covering 8 days of infection had high enough *C. jejuni* isolate counts to be included in this study (as determined by CG8421 genome coverage), as indicated in Sup. Table 1, and were produced by male and female monkeys with a median age of 14 months as noted in Table 1. The *Aotus* inoculum was prepared independently of the human inoculum, and similarly, is not derived from a single colony. This accounts for differences in standing genetic variation between the two inoculums. Therefore, the *Aotus* infection isolate population genomes were compared to the *Aotus* inoculum during genetic analysis of the *Aotus* samples. The samples used in this study were selected from an *Aotus nancymaae* infection preformed as a randomized controlled immunization and *C. jejuni* CG8421 challenge model. Power calculations were used to determine the appropriate number of animals to be used in that preventative treatment investigation model. To be consistent with our human isolate data, we only used samples if: 1) the sample represented symptomatic disease (diarrhea), 2) the sample was produced by an animal that did not receive therapy before or at the time the sample was produced and 3) genomic sequencing yielded sufficient coverage of the *C. jejuni* CG8421 genome for genetic analysis. Therefore, it was necessary that choosing samples for use in our study be performed in a nonblinded manner.

### 5.1.10 *Cj0617/618* Flagellin Mobility Shift Assay

The following genetic studies were performed on strain 81-176 because strain GC8421 is not amenable to genetic analyses. The homopolymeric tract in *Cj0617/618* was repaired as previously described[55]. Briefly, the genes were PCR amplified and cloned behind a sigma28 promoter. The G tract was repaired such that the two genes

were fused into a single open reading frame, and the codons were modified to introduce silent changes that removed the homopolymeric tract by Quick Change mutagenesis using primers pg.16.18 (5-CTCCATTTAAACTAATGAGAGGCGGCGGTATTAGAACGATTTTGTTTGG-3') and pg.16.19 (5'-CCAAACAAAATCGTTCTAATACCGCCGCCTCTCATTAGTTTAAATGGAG-3'). The repaired gene and a kanamycin resistance cassette were introduced into the arylsulfatase gene of 81-176. Purified flagellins were separated using ampholytes ranging from 4-6 (Biolyte 4/6; BioRad) as previously described[209]. IEF protein markers were purchased from Serva.

### 5.1.11 Data Availability

All genomic and RNA sequencing data used in this study have been deposited in the NCBR sequence read archive under Bioproject PRJNA392448.

## 5.2 Experimental Methods Used in Chapter 3

### 5.2.1 ETEC Human Infection Model

Volunteers for Enterotoxigenic E. coli H10407 infection (ClinicalTrials.gov Identifier NCT01922856, Navy Medical Research Center IRB NMRC.2013.0011) were healthy adults from the mid-Atlantic region of the United States. Volunteers signed informed consent forms before participating and allowed for future use of samples produced during the study. Volunteers were either male or female between 18-50 years of age, were in general good health, without significant medical history or previous

exposure to ETEC or *Vibrio cholerae*, did not have an abnormal stool patterns, and did not regularly use laxatives or antacids. On the day of the challenge, volunteers drank 120mL of bicarbonate buffer, followed by 30mL of bicarbonate buffer containing $1.2x10^7$ CFU of virulent ETEC H10407. Inoculum bacteria were initial grown overnight at 37°C on 3 CFA agar plates supplemented with bile salt, and then suspended in PBS. PBS resuspension was kept on ice until being incorporated into the final bicarbonate buffer inoculum given to volunteers. Inoculum concentration was initially estimated via optical density $600_{nm}$ readings on the day of the challenge and then confirmed using colony forming unit enumeration on LB agar plates after overnight incubation at 37°C.

### 5.2.2 Infected Sample Preparation for RNA-Sequencing

ETEC H10407 infected diarrhea was weighed and added 1:1 in RNA later reagent as quickly as possible after production and frozen at -80C. We only worked with infected diarrhea samples from volunteers who were not vaccinated against ETEC, were not receiving antibiotic treatment at the time of production, did not have a restricted diet, and were not receiving prescribed rehydration therapy. Preserved infected samples were thawed in a secondary container on ice and total RNA was extracted via Trizol-chloroform phase separation. DNA was removed (Turbo DNA-free DNAse kit, Ambion) and rRNA was depleted (Ribo-Zero human and bacterial rRNA removal kits, Illumina). Libraries were built for Illumina next-generation sequencing (Ultra Direcitonal RNA Library Prep Kit for Illumina, New England Biolabs).

### 5.2.3 *In vitro* ETEC Sample Preparation for RNA-Sequencing

Enterotoxigenic *E. coli* strain H10407 was grown at 37°C in either standard aerobic atmospheric conditions, or in a in a Coy Laboratories atmosphere controlled chamber for low-oxygen growths. The microaerobic atmosphere used was 85% nitrogen, 10% carbon dioxide, and 5% oxygen, while the anaerobic atmosphere was 85% nitrogen, 10% carbon dioxide, 5% hydrogen. For bacterial growth, CFA agar plates were streaked from a frozen stock of ETEC H10407 and grown overnight at 37°C in the indicated atmosphere. Single colonies were used to inoculate 5 mL aliquots of CFA broth in triplicate and grown with shaking in the indicated atmosphere. For low-oxygen atmosphere growths, CFA broth aliquots were equilibrated with the atmosphere overnight with shaking at 37°C prior to inoculation. For RNA-sequencing, all cultures were grown in triplicate to mid-log phase in the respective atmosphere, harvested via centrifugation within the respective atmosphere, and resuspended in Trizol reagent within the respective atmosphere for immediate RNA extraction. DNA was then removed (Turbo DNA Free DNAse Kit, Ambion), rRNA was removed (Ribo-Zero Bacterial rRNA removal Kit, Illumina) and the samples were prepared for Illumina sequencing using the TruSeq Stranded mRNA Library Kit (Illumina).

### 5.2.4 Illumina RNA-sequencing and Analysis

Infected samples are complex and ETEC RNA represents a small portion of total sample RNA. To identify ETEC-specific transcripts we trimmed sample reads to 21 base pairs in length and then mapped them with 100% exact sequence homology to the ETEC H10407 reference genome using CLC Genomics Workbench software, as has been previously described with similarly complex samples[9]. At least 3 million reads mapped to

the H10407 reference genome per sample. 5 infected diarrhea samples each from different volunteers were sequenced and analyzed individually, and then together as a group represented our *in vivo* condition for analysis. For *in vitro* samples, reads were trimmed and mapped in an identical manner, with at least 15 million reads mapping to the reference genome, and all *in vitro* conditions represent a group of 3 biological replicates.

CLC Genomics WorkBench (Qiagen) was used to to determine differential gene expression fold changess between conditions and a false discovery rate corrected (FDR) p-value was used to determine statistically significance. We considered a significant change in expression between groups to be a weighted proportions fold change > |3| with a FDR p-value of <0.05.

To determine if any native commensal *E. coli* were influencing the mapped reads from *in vivo* samples, we compared the ratio of chromosomally mapped reads vs reads mapped to ETEC-specific virulence plasmids. We assumed commensal *E. coli* genomes would share extensive homology to the ETEC chromosome, and therefore if commensal *E. coli* were influential, chromosomal reads would be over-represented compared to reads mapped to ETEC virulence plasmids. We did not find chromosomal reads to be over-represented, and therefore we did not see a large representation of non-ETEC reads mapping to the ETEC reference genome.

### 5.2.5 ETEC Recombineering and Complementation

To generate a knockout of *fnr* in ETEC H10407 lambda red-mediated gene replacement was used as previously described in *E. coli*[210]. Briefly, the kanamycin cassette from the pkD4 plasmid was PCR amplified with 50 basepair flanking regions which were homologous to the immediate up and downstream regions of the

chromosomally encoded *fnr* gene. This linear PCR product was cleaned using the QIAquick PCR Purification Kit and electroporated into H10407 containing the pkD46 plasmid, which produces the lambda red recombinase, and mediates the homologous recombination of the *fnr* locus with linear PCR product. Recombinant bacteria were selected for on kanamycin LB agar and *fnr*:kan mutants were verified by PCR, sanger sequencing, and by observing an expected growth defect in low-oxygen conditions compared to WT (Supplementary Figure 1). *fnr*:kan was maintained on agar plates and in liquid medium in the presence of kanamycin.

The *fnr* complementation plasmid (pBad18-*fnr*) is an arabinose inducible pBad18 plasmid carrying a chloramphenicol resistance cassette and was built using Gibson assembly. pBad18-*fnr* was electroporated into the *fnr*:kan H10407 genetic background. Complementation was determined by PCR verification of the plasmid, alongside rescue of the *fnr*:kan anaerobic growth defect (Fig. 3.3) as well as rescue of the increased biofilm formation phenotype (Fig. 3.9). *fnr*:kan pBad18-fnr was maintained in the presence of kanamycin and chloramphenicol and with 0.2% arabinose when indicated.

### 5.2.6 Electrophoretic Mobility Shift Assays (EMSA)

FNR activity is sensitive to oxygen, making most standard in vitro assays, including purification, difficult. To address this, we used a well characterized FNR variant, $(FNRD154A)_2$, which is physically linked as an active dimer and is therefore active in aerobic environments. This variant has been validated for use in EMSA assays, and purified $(FNRD154A)_2$-$His_6$ was generously provided to us by Lisa Kay Nolan and

prepared as previously described[198]. Purified protein was stored at -80C in binding buffer (20mM Tris HCL, 50mM NaCl, 40mM EDTA, 4mM DTT, 10% glycerol, pH 6.8).

All assays were carried out as previously described[198] by mixing increasing amounts of $(FNRD154A)_2$-His6 with the PCR amplified promoter regions of virulence factor genes as the DNA probe. PCR amplified promoter regions were size-selected using a Zymoclean DNA Gel Recovery kit (Zymogen) and then cleaned using the DNA Clean and Concentrator kit (Zymogen). Binding reactions of $(FNRD154A)_2$-His6 to DNA probes were performed in binding buffer mentioned above with 0.5mg/ml bovine serum albumin, in 25 uL total volume, at 37°C for 30 minutes. Concentrations of protein and probe used are detailed per assay in supplementary figures 3-6 for clarity. Reaction mixtures were run in a 6% polyacrylamide DNA retardation gel (Invitrogen) in 0.5 TBE buffer (Invitrogen) at 200V for 45 minutes and gels were stained with ethidium bromide in 0.5 TBE buffer for 10 minutes at room temperature before being exposed to ultraviolet light for imaging.

### 5.2.7 Biofilm Production and Measurement

We found ETEC H10407 spontaneously grew biofilms on glass test tubes when grown anaerobically at 37C in 5 mL of CFA broth with vigorous shaking. To quantitate biofilm production, growth media was removed from test tubes using a serological pipette and then 0.1% filter sterilized crystal violet stain was added to the tube to completely submerge the biofilm and incubated for 20 minutes at room temperature. The stain was removed via serological pipette, and then stained biofilms were carefully washed at least 3 times with phosphate buffered saline to be sure there was no residual stain remaining in the tube. Stained and washed biofilms were air dried at room

temperature for 30 minutes and then solubilized in 95% ethanol via pipetting. Dilutions of the solubilized stain were pipetted into a 96-well clear bottom plate, and then the optical density $595_{nm}$ was determined for each sample using 95% ethanol as a blank. All conditions were tested in biological triplicate. Statistical significance was determined across samples using a one-way ANOVA followed by Tukey's multiple comparison test and error bars represent the standard error of the mean.

## References


1. GBD Diarrhoeal Diseases Collaborators. Estimates of global, regional, and national morbidity, mortality, and aetiologies of diarrhoeal diseases: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet Infect. Dis.* **17,** 909–948 (2017).
2. Amour, C. *et al.* Epidemiology and Impact of Campylobacter Infection in Children in 8 Low-Resource Settings: Results From the MAL-ED Study. *Clin. Infect. Dis.* **63,** 1171–1179 (2016).
3. Gupta, S. K. *et al.* Part III. Analysis of data gaps pertaining to enterotoxigenic Escherichia coli infections in low and medium human development index countries, 1984–2005. *Epidemiol. Amp Infect.* **136,** 721–738 (2008).
4. DuPont, H. L. *et al.* Pathogenesis of Escherichia coli Diarrhea. *N. Engl. J. Med.* **285,** 1–9 (1971).
5. Dekeyser, P., Gossuin-Detrain, M., Butzler, J. P. & Sternon, J. Acute Enteritis Due to Related Vibrio: First Positive Stool Cultures. *J. Infect. Dis.* **125,** 390–392 (1972).
6. Lieberman, T. D. *et al.* Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. *Nat. Genet.* **43,** 1275–1280 (2011).
7. Lieberman, T. D. *et al.* Genomic diversity in autopsy samples reveals within-host dissemination of HIV-associated *Mycobacterium tuberculosis*. *Nat. Med.* **22,** 1470–1474 (2016).
8. Viberg, L. T. *et al.* Within-Host Evolution of Burkholderia pseudomallei during Chronic Infection of Seven Australasian Cystic Fibrosis Patients. *mBio* **8,** (2017).

9. Crofts, A. A. *et al*. Campylobacter jejuni transcriptional and genetic adaptation during human infection. *Nat. Microbiol.* **3,** 494–502 (2018).

10. Scallan, E. *et al*. Foodborne illness acquired in the United States--major pathogens. *Emerg. Infect. Dis.* **17,** 7–15 (2011).

11. Lahti, E., Löfdahl, M., Ågren, J., Hansson, I. & Olsson Engvall, E. Confirmation of a Campylobacteriosis Outbreak Associated with Chicken Liver Pâté Using PFGE and WGS. *Zoonoses Public Health* **64,** 14–20 (2017).

12. Costard, S., Espejo, L., Groenendaal, H. & Zagmutt, F. J. Outbreak-Related Disease Burden Associated with Consumption of Unpasteurized Cow's Milk and Cheese, United States, 2009-2014. *Emerg. Infect. Dis.* **23,** 957–964 (2017).

13. Lee, G. *et al*. Symptomatic and Asymptomatic Campylobacter Infections Associated with Reduced Growth in Peruvian Children. *PLoS Negl. Trop. Dis.* **7,** (2013).

14. Münchener medizinische Wochenschrift. *Issues 17 Oct 1899- Have TitleMünchener Med. Wochenschr.* volumes (1886).

15. Butzler, J. P., De Boeck, M. & Goossens, H. New selective medium for isolation of Campylobacter jejuni from faecal specimens. *Lancet Lond. Engl.* **1,** 818 (1983).

16. Endtz, H. P. *et al*. Comparison of six media, including a semisolid agar, for the isolation of various Campylobacter species from stool specimens. *J. Clin. Microbiol.* **29,** 1007–1010 (1991).

17. Rodrigues, R. C. *et al*. Description of Campylobacter jejuni Bf, an atypical aero-tolerant strain. *Gut Pathog.* **7,** 30 (2015).

18. Skuhala, T., Škerk, V., Markotić, A., Bukovski, S. & Desnica, B. Septic abortion caused by Campylobacter jejuni bacteraemia. *J. Chemother. Florence Italy* **28,** 335–336 (2016).

19. Hannu, T. *et al*. Campylobacter-triggered reactive arthritis: a population-based study. *Rheumatol. Oxf. Engl.* **41,** 312–318 (2002).

20. Kaldor, J. & Speed, B. R. Guillain-Barré syndrome and Campylobacter jejuni: a serological study. *Br. Med. J. Clin. Res. Ed* **288,** 1867–1870 (1984).

21. Yuki, N., Yoshino, H., Sato, S. & Miyatake, T. Acute axonal polyneuropathy associated with anti-GM1 antibodies following Campylobacter enteritis. *Neurology* **40,** 1900–1902 (1990).

22. Bech, E., Orntoft, T. F., Andersen, L. P., Skinhøj, P. & Jakobsen, J. IgM anti-GM1 antibodies in the Guillain-Barré syndrome: a serological predictor of the clinical course. *J. Neuroimmunol.* **72,** 59–66 (1997).

23. Karlyshev, A. V., Linton, D., Gregson, N. A., Lastovica, A. J. & Wren, B. W. Genetic and biochemical evidence of a Campylobacter jejuni capsular polysaccharide that accounts for Penner serotype specificity. *Mol. Microbiol.* **35,** 529–541 (2000).

24. Tribble, D. R. *et al*. Campylobacter jejuni strain CG8421: a refined model for the study of Campylobacteriosis and evaluation of Campylobacter vaccines in human subjects. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* **49,** 1512–1519 (2009).

25. Parkhill, J. *et al*. The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hypervariable sequences. *Nature* **403,** 665–668 (2000).

26. Bayliss, C. D. *et al*. Phase variable genes of Campylobacter jejuni exhibit high mutation rates and specific mutational patterns but mutability is not the major determinant of population structure during host colonization. *Nucleic Acids Res*. **40,** 5876–5889 (2012).

27. Poly, F. & Guerry, P. Pathogenesis of Campylobacter. *Curr*. *Opin*. *Gastroenterol*. **24,** 27–31 (2008).

28. Grant, C. C., Konkel, M. E., Cieplak, W. & Tompkins, L. S. Role of flagella in adherence, internalization, and translocation of Campylobacter jejuni in nonpolarized and polarized epithelial cell cultures. *Infect*. *Immun*. **61,** 1764–1771 (1993).

29. Black, R. E., Levine, M. M., Clements, M. L., Hughes, T. P. & Blaser, M. J. Experimental Campylobacter jejuni infection in humans. *J*. *Infect*. *Dis*. **157,** 472–479 (1988).

30. Jin, S., Song, Y. C., Emili, A., Sherman, P. M. & Chan, V. L. JlpA of Campylobacter jejuni interacts with surface-exposed heat shock protein 90alpha and triggers signalling pathways leading to the activation of NF-kappaB and p38 MAP kinase in epithelial cells. *Cell*. *Microbiol*. **5,** 165–174 (2003).

31. Jin, S. *et al*. JlpA, a novel surface-exposed lipoprotein specific to Campylobacter jejuni, mediates adherence to host epithelial cells. *Mol*. *Microbiol*. **39,** 1225–1236 (2001).

32. Konkel, M. E., Garvis, S. G., Tipton, S. L., Anderson, D. E. & Cieplak, W. Identification and molecular cloning of a gene encoding a fibronectin-binding protein (CadF) from Campylobacter jejuni. *Mol. Microbiol.* **24,** 953–963 (1997).

33. Konkel, M. E., Kim, B. J., Rivera-Amill, V. & Garvis, S. G. Bacterial secreted proteins are required for the internaliztion of Campylobacter jejuni into cultured mammalian cells. *Mol. Microbiol.* **32,** 691–701 (1999).

34. Poli, V. F. S., Thorsen, L., Olesen, I., Wik, M. T. & Jespersen, L. Differentiation of the virulence potential of Campylobacter jejuni strains by use of gene transcription analysis and a Caco-2 assay. *Int. J. Food Microbiol.* **155,** 60–68 (2012).

35. Konkel, M. E. *et al*. Secretion of virulence proteins from Campylobacter jejuni is dependent on a functional flagellar export apparatus. *J. Bacteriol.* **186,** 3296–3303 (2004).

36. Barrero-Tobon, A. M. & Hendrixson, D. R. Identification and analysis of flagellar coexpressed determinants (Feds) of Campylobacter jejuni involved in colonization. *Mol. Microbiol.* **84,** 352–369 (2012).

37. Jinadasa, R. N., Bloom, S. E., Weiss, R. S. & Duhamel, G. E. Cytolethal distending toxin: a conserved bacterial genotoxin that blocks cell cycle progression, leading to apoptosis of a broad range of mammalian cell lineages. *Microbiol. Read. Engl.* **157,** 1851–1875 (2011).

38. Shenker, B. J. *et al*. Induction of cell cycle arrest in lymphocytes by Actinobacillus actinomycetemcomitans cytolethal distending toxin requires three subunits for maximum activity. *J. Immunol. Baltim. Md 1950* **174,** 2228–2234 (2005).

39. Hickey, T. E. *et al*. Campylobacter jejuni cytolethal distending toxin mediates release of interleukin-8 from intestinal epithelial cells. *Infect. Immun.* **68,** 6535–6541 (2000).

40. Abuoun, M. *et al*. Cytolethal distending toxin (CDT)-negative Campylobacter jejuni strains and anti-CDT neutralizing antibodies are induced during human infection but not during colonization in chickens. *Infect. Immun.* **73,** 3053–3062 (2005).

41. Mortensen, N. P. *et al*. The role of Campylobacter jejuni cytolethal distending toxin in gastroenteritis: toxin detection, antibody production, and clinical outcome. *APMIS Acta Pathol. Microbiol. Immunol. Scand.* **119,** 626–634 (2011).

42. Lara-Tejero, M. & Galán, J. E. Cytolethal distending toxin: limited damage as a strategy to modulate cellular functions. *Trends Microbiol.* **10,** 147–152 (2002).

43. Nakashige, T. G., Zhang, B., Krebs, C. & Nolan, E. M. Human calprotectin is an iron-sequestering host-defense protein. *Nat. Chem. Biol.* **11,** 765–771 (2015).

44. Arezes, J. *et al*. Hepcidin-induced hypoferremia is a critical host defense mechanism against the siderophilic bacterium Vibrio vulnificus. *Cell Host Microbe* **17,** 47–57 (2015).

45. Andrews, S. C., Robinson, A. K. & Rodríguez-Quiñones, F. Bacterial iron homeostasis. *FEMS Microbiol. Rev.* **27,** 215–237 (2003).

46. Cox, C. D. Iron uptake with ferripyochelin and ferric citrate by Pseudomonas aeruginosa. *J. Bacteriol.* **142,** 581–587 (1980).

47. Moore, D. G. & Earhart, C. F. Specific inhibition of Escherichia coli ferrienterochelin uptake by a normal human serum immunoglobulin. *Infect. Immun.* **31,** 631–635 (1981).

48. Knüsel, F. & Nüesch, J. Mechanism of action of sideromycins. *Nature* **206,** 674–676 (1965).

49. Anderson, B. F. *et al*. Models for the bacterial iron-transport chelate enterochelin. *Nature* **262,** 722–724 (1976).

50. Naikare, H. *et al*. Campylobacter jejuni ferric-enterobactin receptor CfrA is TonB3 dependent and mediates iron acquisition from structurally different catechol siderophores. *Met. Integr. Biometal Sci.* **5,** 988–996 (2013).

51. Bergman, M., Del Prete, G., van Kooyk, Y. & Appelmelk, B. Helicobacter pylori phase variation, immune modulation and gastric autoimmunity. *Nat. Rev. Microbiol.* **4,** 151–159 (2006).

52. Kim, J.-S. *et al*. Passage of Campylobacter jejuni through the chicken reservoir or mice promotes phase variation in contingency genes Cj0045 and Cj0170 that strongly associates with colonization and disease in a mouse model. *Microbiology* **158,** 1304–1316 (2012).

53. Bayliss, C. D. *et al*. Phase variable genes of Campylobacter jejuni exhibit high mutation rates and specific mutational patterns but mutability is not the major determinant of population structure during host colonization. *Nucleic Acids Res.* **40,** 5876–5889 (2012).

54. Cameron, A. *et al*. High-Frequency Variation of Purine Biosynthesis Genes Is a Mechanism of Success in Campylobacter jejuni. *mBio* **6,** e00612-15 (2015).

55. Pequegnat, B. *et al*. Phase variable changes in the position of O-methyl phosphoramidate modifications on the polysaccharide capsule of Campylobacter jejuni modulate serum resistance. *J. Bacteriol*. JB.00027-17 (2017). doi:10.1128/JB.00027-17

56. Blaser, M. J. Campylobacter Enteritis in the United States: A Multicenter Study. *Ann. Intern. Med*. **98,** 360 (1983).

57. Ternhag, A., Asikainen, T., Giesecke, J. & Ekdahl, K. A Meta-Analysis on the Effects of Antibiotic Treatment on Duration of Symptoms Caused by Infection with Campylobacter Species. *Clin. Infect. Dis*. **44,** 696–700 (2007).

58. Butzler, J. P., Dekeyser, P. & Lafontaine, T. Susceptibility of Related Vibrios and Vibrio Fetus to Twelve Antibiotics. *Antimicrob. Agents Chemother*. **5,** 86–89 (1974).

59. Nolan, C. M., Johnson, K. E., Coyle, M. B. & Faler, K. Campylobacter jejuni enteritis: efficacy of antimicrobial and antimotility drugs. *Am. J. Gastroenterol*. **78,** 621–626 (1983).

60. Taylor, D. N. *et al*. Rifaximin, a nonabsorbed oral antibiotic, prevents shigellosis after experimental challenge. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am*. **42,** 1283–1288 (2006).

61. Rimmer, J. E. *et al*. Rifaximin Fails to Prevent Campylobacteriosis in the Human Challenge Model: A Randomized, Double-Blind, Placebo-Controlled Trial. *Clin. Infect. Dis*. doi:10.1093/cid/cix1014

62. Monteiro, M. A. *et al*. Capsule Polysaccharide Conjugate Vaccine against Diarrheal Disease Caused by Campylobacter jejuni. *Infect. Immun*. **77,** 1128–1136 (2009).

63. Wildfeuer, A., Laufen, H., Müller-Wening, D. & Haferkamp, O. Interaction of azithromycin and human phagocytic cells. Uptake of the antibiotic and the effect on the survival of ingested bacteria in phagocytes. *Arzneimittelforschung*. **39,** 755–758 (1989).

64. Riddle, M. S. & Guerry, P. Status of vaccine research and development for Campylobacter jejuni. *Vaccine* **34,** 2903–2906 (2016).

65. Maue, A. C., Poly, F. & Guerry, P. A capsule conjugate vaccine approach to prevent diarrheal disease caused by Campylobacter jejuni. *Hum. Vaccines Immunother*. **10,** 1499–1504 (2014).

66. Bacon, D. J. *et al*. A phase-variable capsule is involved in virulence of Campylobacter jejuni 81-176. *Mol. Microbiol*. **40,** 769–777 (2001).

67. Hosangadi, D., Smith, P. G. & Giersing, B. K. Considerations for using ETEC and Shigella disease burden estimates to guide vaccine development strategy. *Vaccine* (2017). doi:10.1016/j.vaccine.2017.09.083

68. Kutas, F., Vetési, F. & Semjén, G. Experimental diarrhoea in baby rabbits due to oral administration of heat-labile enterotoxin of E. coli enteropathogenic for swine. *Acta Vet. Acad. Sci. Hung*. **24,** 177–186 (1974).

69. Donta, S. T., Moon, H. W. & Whipp, S. C. Detection of heat-labile Escherichia coli enterotoxin with the use of adrenal cells in tissue culture. *Science* **183,** 334–336 (1974).

70. Finkelstein, R. A. *et al*. Isolation and properties of heat-labile enterotoxin(s) from enterotoxigenic Escherichia coli. *J. Infect. Dis.* **133 Suppl,** 120–137 (1976).

71. Sack, R. B. *et al*. Enterotoxigenic Escherichia coli isolated from patients with severe cholera-like disease. *J. Infect. Dis.* **123,** 378–385 (1971).

72. Murray, C. J. L. *et al*. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet Lond. Engl.* **380,** 2197–2223 (2012).

73. Schultsz, C. *et al*. Diarrheagenic Escherichia coli and acute and persistent diarrhea in returned travelers. *J. Clin. Microbiol.* **38,** 3550–3554 (2000).

74. Yoder, J. S. *et al*. Outbreak of enterotoxigenic Escherichia coli infection with an unusually long duration of illness. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* **42,** 1513–1517 (2006).

75. Qadri, F., Svennerholm, A.-M., Faruque, A. S. G. & Sack, R. B. Enterotoxigenic Escherichia coli in developing countries: epidemiology, microbiology, clinical features, treatment, and prevention. *Clin. Microbiol. Rev.* **18,** 465–483 (2005).

76. Beatty, M. E. *et al*. Epidemic diarrhea due to enterotoxigenic Escherichia coli. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* **42,** 329–334 (2006).

77. von Sonnenburg, F. *et al*. Risk and aetiology of diarrhoea at various tourist destinations. *Lancet Lond. Engl.* **356,** 133–134 (2000).

78. Haberberger, R. L. *et al*. Travelers' diarrhea among United States military personnel during joint American-Egyptian armed forces exercises in Cairo, Egypt. *Mil. Med.* **156,** 27–30 (1991).

79. Nada, R. A. *et al*. Phenotypic and genotypic characterization of enterotoxigenic Escherichia coli isolated from U.S. military personnel participating in Operation Bright Star, Egypt, from 2005 to 2009. *Diagn. Microbiol. Infect. Dis.* **76,** 272–277 (2013).

80. Harro, C. *et al*. Refinement of a Human Challenge Model for Evaluation of Enterotoxigenic Escherichia coli Vaccines. *Clin. Vaccine Immunol.* **18,** 1719–1727 (2011).

81. Maloo, A., Fulke, A. B., Mulani, N., Sukumaran, S. & Ram, A. Pathogenic multiple antimicrobial resistant Escherichia coli serotypes in recreational waters of Mumbai, India: a potential public health risk. *Environ. Sci. Pollut. Res. Int.* **24,** 11504–11517 (2017).

82. Ram, S., Vajpayee, P. & Shanker, R. Rapid culture-independent quantitative detection of enterotoxigenic Escherichia coli in surface waters by real-time PCR with molecular beacon. *Environ. Sci. Technol.* **42,** 4577–4582 (2008).

83. Sidhu, J. P. S., Ahmed, W., Hodgers, L. & Toze, S. Occurrence of virulence genes associated with Diarrheagenic pathotypes in Escherichia coli isolates from surface water. *Appl. Environ. Microbiol.* **79,** 328–335 (2013).

84. Ahmed, D. *et al*. Presence of enterotoxigenic Escherichia coli in biofilms formed in water containers in poor households coincides with epidemic seasons in Dhaka. *J. Appl. Microbiol*. **114,** 1223–1229 (2013).

85. Evans, D. G., Silver, R. P., Evans, D. J., Chase, D. G. & Gorbach, S. L. Plasmid-controlled colonization factor associated with virulence in Esherichia coli enterotoxigenic for humans. *Infect. Immun*. **12,** 656–667 (1975).

86. Levine, M. M. *et al*. Diarrhea caused by Escherichia coli that produce only heat-stable enterotoxin. *Infect. Immun*. **17,** 78–82 (1977).

87. Li, Y.-F. *et al*. Structure of CFA/I fimbriae from enterotoxigenic Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A*. **106,** 10793–10798 (2009).

88. Rönnberg, B., Wadström, T. & Jörnvall, H. Structure of a heat-stable enterotoxin produced by a human strain of Escherichia coli. Differences from the toxin of another human strain suggest the presence of compensated amino acid exchanges. *FEBS Lett*. **155,** 183–186 (1983).

89. Yamamoto, T., Tamura, T. & Yokota, T. Primary structure of heat-labile enterotoxin produced by Escherichia coli pathogenic for humans. *J. Biol. Chem*. **259,** 5037–5044 (1984).

90. von Mentzer, A. *et al*. Identification of enterotoxigenic Escherichia coli (ETEC) clades with long-term global distribution. *Nat. Genet*. **46,** 1321–1326 (2014).

91. Sahl, J. W. *et al*. Examination of the Enterotoxigenic Escherichia coli Population Structure during Human Infection. *mBio* **6,** e00501-15 (2015).

92. Yamamoto, T. & Yokota, T. Plasmids of enterotoxigenic Escherichia coli H10407: evidence for two heat-stable enterotoxin genes and a conjugal transfer system. *J. Bacteriol.* **153,** 1352–1360 (1983).

93. Crossman, L. C. *et al.* A commensal gone bad: complete genome sequence of the prototypical enterotoxigenic Escherichia coli strain H10407. *J. Bacteriol.* **192,** 5822–5831 (2010).

94. Canto, F. D. *et al.* Identification of Coli Surface Antigen 23, a Novel Adhesin of Enterotoxigenic Escherichia coli. *Infect. Immun.* **80,** 2791–2801 (2012).

95. Bao, R., Liu, Y., Savarino, S. J. & Xia, D. Off-pathway assembly of fimbria subunits is prevented by chaperone CfaA of CFA/I fimbriae from enterotoxigenic E. coli. *Mol. Microbiol.* **102,** 975–991 (2016).

96. Jones, G. W. & Rutter, J. M. Contribution of the K88 antigen of Escherichia coli to enteropathogenicity; protection against disease by neutralizing the adhesive properties of K88 antigen. *Am. J. Clin. Nutr.* **27,** 1441–1449 (1974).

97. Grange, P. A., Erickson, A. K., Anderson, T. J. & Francis, D. H. Characterization of the carbohydrate moiety of intestinal mucin-type sialoglycoprotein receptors for the K88ac fimbrial adhesin of Escherichia coli. *Infect. Immun.* **66,** 1613–1621 (1998).

98. Francis, D. H. *et al.* Expression of Mucin-Type Glycoprotein K88 Receptors Strongly Correlates with Piglet Susceptibility to K88+ Enterotoxigenic Escherichia coli, but Adhesion of This Bacterium to Brush Borders Does Not. *Infect. Immun.* **66,** 4050–4055 (1998).

99. Zhang, C., Rausch, D. & Zhang, W. Little Heterogeneity among Genes Encoding Heat-Labile and Heat-Stable Toxins of Enterotoxigenic Escherichia coli Strains Isolated from Diarrheal Pigs. *Appl. Environ. Microbiol.* **75,** 6402–6405 (2009).

100. Hodson, C. *et al.* Control of Virulence Gene Expression by the Master Regulator, CfaD, in the Prototypical EnterotoxigenicEscherichia coliStrain, H10407. *Front. Microbiol.* **8,** 1525 (2017).

101. Willshaw, G. A., McConnell, M. M., Smith, H. R. & Rowe, B. Structural and regulatory genes for coli surface associated antigen 4 (CS4) are encoded by separate plasmids in enterotoxigenic Escherichia coli strains of serotype 0.25.H42. *FEMS Microbiol. Lett.* **56,** 255–260 (1990).

102. Jordi, B. J. *et al.* The positive regulator CfaD overcomes the repression mediated by histone-like protein H-NS (H1) in the CFA/I fimbrial operon of Escherichia coli. *EMBO J.* **11,** 2627–2632 (1992).

103. Jordi, B. J. The mode of action of CfaD of Escherichia coli and VirF of Shigella flexneri and other members of the AraC family of positive regulators. *Mol. Microbiol.* **6,** 3451 (1992).

104. Santiago, A. E. *et al.* A large family of antivirulence regulators modulates the effects of transcriptional activators in Gram-negative pathogenic bacteria. *PLoS Pathog.* **10,** e1004153 (2014).

105. Czuprynski, C. J. & Faith, N. G. Sodium Bicarbonate Enhances the Severity of Infection in Neutropenic Mice Orally Inoculated with Listeria monocytogenes EGD. *Clin. Diagn. Lab. Immunol.* **9,** 477–481 (2002).

106. Effects of Bicarbonate on Growth of Neisseria gonorrhoeae: Replacement of Gaseous CO2 Atmosphere. *Appl. Microbiol.* **30,** 162 (1975).

107. Pilonieta, M. C., Bodero, M. D. & Munson, G. P. CfaD-dependent expression of a novel extracytoplasmic protein from enterotoxigenic Escherichia coli. *J. Bacteriol.* **189,** 5060–5067 (2007).

108. Taxt, A., Aasland, R., Sommerfelt, H., Nataro, J. & Puntervoll, P. Heat-Stable Enterotoxin of Enterotoxigenic Escherichia coli as a Vaccine Target. *Infect. Immun.* **78,** 1824–1831 (2010).

109. Rasheed, J. K., Guzmán-Verduzco, L. M. & Kupersztoch, Y. M. Two precursors of the heat-stable enterotoxin of Escherichia coli: evidence of extracellular processing. *Mol. Microbiol.* **4,** 265–273 (1990).

110. Ozaki, H. *et al*. Molecular structure of the toxin domain of heat-stable enterotoxin produced by a pathogenic strain of Escherichia coli. A putative binding site for a binding protein on rat intestinal epithelial cell membranes. *J. Biol. Chem.* **266,** 5934–5941 (1991).

111. Schulz, S., Green, C. K., Yuen, P. S. & Garbers, D. L. Guanylyl cyclase is a heat-stable enterotoxin receptor. *Cell* **63,** 941–948 (1990).

112. Field, M., Graf, L. H., Laird, W. J. & Smith, P. L. Heat-stable enterotoxin of Escherichia coli: in vitro effects on guanylate cyclase activity, cyclic GMP concentration, and ion transport in small intestine. *Proc. Natl. Acad. Sci. U. S. A.* **75,** 2800–2804 (1978).

113. Vaandrager, A. B. *et al*. Membrane targeting of cGMP-dependent protein kinase is required for cystic fibrosis transmembrane conductance regulator Cl- channel activation. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 1466–1471 (1998).

114. Bakre, M. M. & Visweswariah, S. S. Dual regulation of heat-stable enterotoxin-mediated cGMP accumulation in T84 cells by receptor desensitization and increased phosphodiesterase activity. *FEBS Lett.* **408,** 345–349 (1997).

115. Chao, A. C. *et al*. Activation of intestinal CFTR Cl- channel by heat-stable enterotoxin and guanylin via cAMP-dependent protein kinase. *EMBO J.* **13,** 1065–1072 (1994).

116. Honegger, K. J. *et al*. Regulation of sodium-proton exchanger isoform 3 (NHE3) by PKA and exchange protein directly activated by cAMP (EPAC). *Proc. Natl. Acad. Sci.* **103,** 803–808 (2006).

117. Schulz, S., Lopez, M. J., Kuhn, M. & Garbers, D. L. Disruption of the guanylyl cyclase-C gene leads to a paradoxical phenotype of viable but heat-stable enterotoxin-resistant mice. *J. Clin. Invest.* **100,** 1590–1595 (1997).

118. Yamamoto, T., Nakazawa, T., Miyata, T., Kaji, A. & Yokota, T. Evolution and structure of two ADP-ribosylation enterotoxins, Escherichia coli heat-labile toxin and cholera toxin. *FEBS Lett.* **169,** 241–246 (1984).

119. Hofstra, H. & Witholt, B. Heat-labile enterotoxin in Escherichia coli. Kinetics of association of subunits into periplasmic holotoxin. *J. Biol. Chem.* **260,** 16037–16044 (1985).

120. Tauschek, M., Gorrell, R. J., Strugnell, R. A. & Robins-Browne, R. M. Identification of a protein secretory pathway for the secretion of heat-labile enterotoxin by an enterotoxigenic strain of Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 7066–7071 (2002).

121. Horstman, A. L. & Kuehn, M. J. Enterotoxigenic Escherichia coli secretes active heat-labile enterotoxin via outer membrane vesicles. *J. Biol. Chem.* **275,** 12489–12496 (2000).

122. Moss, J., Osborne, J. C., Fishman, P. H., Nakaya, S. & Robertson, D. C. Escherichia coli heat-labile enterotoxin. Ganglioside specificity and ADP-ribosyltransferase activity. *J. Biol. Chem.* **256,** 12861–12865 (1981).

123. Lencer, W. I., Hirst, T. R. & Holmes, R. K. Membrane traffic and the cellular uptake of cholera toxin. *Biochim. Biophys. Acta BBA - Mol. Cell Res.* **1450,** 177–190 (1999).

124. Roy, K., Hamilton, D. J., Munson, G. P. & Fleckenstein, J. M. Outer Membrane Vesicles Induce Immune Responses to Virulence Proteins and Protect against Colonization by Enterotoxigenic Escherichia coli. *Clin. Vaccine Immunol.* **18,** 1803–1808 (2011).

125. Lääveri, T., Vilkman, K., Pakkanen, S., Kirveskari, J. & Kantele, A. Despite antibiotic treatment of travellers' diarrhoea pathogens are found in stools from half of travellers at return. *Travel Med. Infect. Dis.* (2018). doi:10.1016/j.tmaid.2018.04.003

126. Chakraborty, S. *et al*. Impact of lower challenge doses of enterotoxigenic Escherichia coli on clinical outcome, intestinal colonization and immune responses in adult volunteers. *PLoS Negl. Trop. Dis*. **12,** e0006442 (2018).

127. Hu, Y., Ren, J., Zhan, M., Li, W. & Dai, H. Efficacy of rifaximin in prevention of travelers' diarrhea: a meta-analysis of randomized, double-blind, placebo-controlled trials. *J. Travel Med*. **19,** 352–356 (2012).

128. Johnson, P. C. *et al*. Comparison of loperamide with bismuth subsalicylate for the treatment of acute travelers' diarrhea. *JAMA* **255,** 757–760 (1986).

129. Haycocks, J. R. J., Sharma, P., Stringer, A. M., Wade, J. T. & Grainger, D. C. The Molecular Basis for Control of ETEC Enterotoxin Expression in Response to Environment and Host. *PLOS Pathog*. **11,** e1004605 (2015).

130. Guo, B. *et al*. CmeR Functions as a Pleiotropic Regulator and Is Required for Optimal Colonization of Campylobacter jejuni In Vivo. *J. Bacteriol*. **190,** 1879–1890 (2008).

131. Hoffmann, A. A. & Willi, Y. Detecting genetic responses to environmental change. *Nat. Rev. Genet*. **9,** 421–432 (2008).

132. Fields, J. A. & Thompson, S. A. Campylobacter jejuni CsrA Mediates Oxidative Stress Responses, Biofilm Formation, and Host Cell Invasion. *J. Bacteriol*. **190,** 3411–3416 (2008).

133. Lin, J., Sahin, O., Michel, L. O. & Zhang, Q. Critical Role of Multidrug Efflux Pump CmeABC in Bile Resistance and In Vivo Colonization of Campylobacter jejuni. *Infect. Immun*. **71,** 4250–4259 (2003).

134. Lei, H.-T. *et al*. Crystal structures of CmeR-bile acid complexes from Campylobacter jejuni. *Protein Sci. Publ. Protein Soc.* **20,** 712–723 (2011).

135. Subashchandrabose, S. *et al*. Host-specific induction of Escherichia coli fitness genes during human urinary tract infection. *Proc. Natl. Acad. Sci.* **111,** 18327–18332 (2014).

136. Velculescu, V. E. *et al*. Characterization of the yeast transcriptome. *Cell* **88,** 243–251 (1997).

137. Chomczynski, P. & Sacchi, N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162,** 156–159 (1987).

138. Aronson, A. I. & McCARTHY, B. J. Studies of E. coli ribosomal RNA and its degradation products. *Biophys. J.* **1,** 215–226 (1961).

139. He, S. *et al*. Validation of two ribosomal RNA removal methods for microbial metatranscriptomics. *Nat. Methods* **7,** 807–812 (2010).

140. Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M. & Gilad, Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.* **18,** 1509–1517 (2008).

141. Bullard, J. H., Purdom, E., Hansen, K. D. & Dudoit, S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics* **11,** 94 (2010).

142. Baggerly, K. A., Deng, L., Morris, J. S. & Aldaz, C. M. Differential expression in SAGE: accounting for normal between-library variation. *Bioinforma. Oxf. Engl.* **19,** 1477–1483 (2003).

143. Goodall, E. C. A. *et al.* The Essential Genome of Escherichia coli K-12. *mBio* **9,** e02096-17 (2018).

144. Thomas, D. K. *et al.* Comparative Variation within the Genome of Campylobacter jejuni NCTC 11168 in Human and Murine Hosts. *PLOS ONE* **9,** e88229 (2014).

145. Yang, L. *et al.* Evolutionary dynamics of bacteria in a human host environment. *Proc. Natl. Acad. Sci.* **108,** 7481–7486 (2011).

146. Kennemann, L. *et al.* Helicobacter pylori genome evolution during human infection. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 5033–5038 (2011).

147. Harris, S. R. *et al.* Evolution of MRSA During Hospital Transmission and Intercontinental Spread. *Science* **327,** 469–474 (2010).

148. Zdziarski, J. *et al.* Host Imprints on Bacterial Genomes—Rapid, Divergent Evolution in Individual Patients. *PLOS Pathog.* **6,** e1001078 (2010).

149. Copin, R. *et al.* Within Host Evolution Selects for a Dominant Genotype of Mycobacterium tuberculosis while T Cells Increase Pathogen Genetic Diversity. *PLOS Pathog.* **12,** e1006111 (2016).

150. Srikhanta, Y. N., Fox, K. L. & Jennings, M. P. The phasevarion: phase variation of type III DNA methyltransferases controls coordinated switching in multiple genes. *Nat. Rev. Microbiol.* **8,** 196–206 (2010).

151. Jerome, J. P. *et al*. Standing Genetic Variation in Contingency Loci Drives the Rapid Adaptation of Campylobacter jejuni to a Novel Host. *PLOS ONE* **6,** e16399 (2011).

152. Mohawk, K. L., Poly, F., Sahl, J. W., Rasko, D. A. & Guerry, P. High Frequency, Spontaneous motA Mutations in Campylobacter jejuni Strain 81-176. *PLOS ONE* **9,** e88043 (2014).

153. Cawthraw, S. A., Wassenaar, T. M., Ayling, R. & Newell, D. G. Increased colonization potential of Campylobacter jejuni strain 81116 after passage through chickens and its implication on the rate of transmission within flocks. *Epidemiol. Infect*. **117,** 213–215 (1996).

154. Baqar, S. *et al*. Recrudescent Campylobacter jejuni infection in an immunocompetent adult following experimental infection with a well-characterized organism. *Clin. Vaccine Immunol. CVI* **17,** 80–86 (2010).

155. Palyada, K., Threadgill, D. & Stintzi, A. Iron Acquisition and Regulation in Campylobacter jejuni. *J. Bacteriol*. **186,** 4714–4729 (2004).

156. Brás, A. M., Chatterjee, S., Wren, B. W., Newell, D. G. & Ketley, J. M. A Novel Campylobacter jejuniTwo-Component Regulatory System Important for Temperature-Dependent Growth and Colonization. *J. Bacteriol*. **181,** 3298–3302 (1999).

157. Del Rocio Leon-Kempis, M., Guccione, E., Mulholland, F., Williamson, M. P. & Kelly, D. J. The Campylobacter jejuni PEB1a adhesin is an aspartate/glutamate-

binding protein of an ABC transporter essential for microaerobic growth on dicarboxylic amino acids. *Mol. Microbiol.* **60,** 1262–1275 (2006).

158. Ziprin, R. L. *et al.* Role of Campylobacter jejuni potential virulence genes in cecal colonization. *Avian Dis.* **45,** 549–557 (2001).

159. Jones, M. A. *et al.* Adaptation of Campylobacter jejuni NCTC11168 to High-Level Colonization of the Avian Gastrointestinal Tract. *Infect. Immun.* **72,** 3769–3776 (2004).

160. Taveirne, M. E., Theriot, C. M., Livny, J. & DiRita, V. J. The Complete Campylobacter jejuni Transcriptome during Colonization of a Natural Host Determined by RNAseq. *PLOS ONE* **8,** e73586 (2013).

161. Corcionivoschi, N. *et al.* Mucosal reactive oxygen species decrease virulence by disrupting Campylobacter jejuni phosphotyrosine signaling. *Cell Host Microbe* **12,** 47–59 (2012).

162. Pryjma, M., Apel, D., Huynh, S., Parker, C. T. & Gaynor, E. C. FdhTU-Modulated Formate Dehydrogenase Expression and Electron Donor Availability Enhance Recovery of Campylobacter jejuni following Host Cell Infection. *J. Bacteriol.* **194,** 3803–3813 (2012).

163. Johnson, J. G., Gaddy, J. A. & DiRita, V. J. The PAS Domain-Containing Protein HeuR Regulates Heme Uptake in Campylobacter jejuni. *mBio* **7,** e01691-16 (2016).

164. Hoang, K. V., Wang, Y. & Lin, J. Identification of genetic loci that contribute to Campylobacter resistance to fowlicidin-1, a chicken host defense peptide. *Front. Cell. Infect. Microbiol.* **2,** (2012).

165. Luethy, P. M., Huynh, S., Parker, C. T. & Hendrixson, D. R. Analysis of the Activity and Regulon of the Two-Component Regulatory System Composed by Cjj81176_1484 and Cjj81176_1483 of Campylobacter jejuni. *J. Bacteriol.* **197,** 1592–1605 (2015).

166. Javed, M. A. *et al.* Transposon mutagenesis in a hyper-invasive clinical isolate of Campylobacter jejuni reveals a number of genes with potential roles in invasion. *Microbiology* **156,** 1134–1143 (2010).

167. Lin, J., Overbye Michel, L. & Zhang, Q. CmeABC Functions as a Multidrug Efflux System in Campylobacter jejuni. *Antimicrob. Agents Chemother.* **46,** 2124–2131 (2002).

168. Grinnage-Pulley, T. & Zhang, Q. Genetic Basis and Functional Consequences of Differential Expression of the CmeABC Efflux Pump in Campylobacter jejuni Isolates. *PLOS ONE* **10,** e0131534 (2015).

169. Guerry, P. *et al.* Changes in flagellin glycosylation affect Campylobacter autoagglutination and virulence. *Mol. Microbiol.* **60,** 299–311 (2006).

170. Howard, S. L. *et al.* Campylobacter jejuni glycosylation island important in cell charge, legionaminic acid biosynthesis, and colonization of chickens. *Infect. Immun.* **77,** 2544–2556 (2009).

171. Murray, J. L., Kwon, T., Marcotte, E. M. & Whiteley, M. Intrinsic Antimicrobial Resistance Determinants in the Superbug Pseudomonas aeruginosa. *mBio* **6,** e01603-1615 (2015).

172. Turner, K. H., Wessel, A. K., Palmer, G. C., Murray, J. L. & Whiteley, M. Essential genome of Pseudomonas aeruginosa in cystic fibrosis sputum. *Proc. Natl. Acad. Sci. U. S. A.* **112,** 4110–4115 (2015).

173. Hodson, C. *et al*. Control of Virulence Gene Expression by the Master Regulator, CfaD, in the Prototypical EnterotoxigenicEscherichia coliStrain, H10407. *Front. Microbiol.* **8,** 1525 (2017).

174. Yamamoto, T., Tamura, T. & Yokota, T. Primary structure of heat-labile enterotoxin produced by Escherichia coli pathogenic for humans. *J. Biol. Chem.* **259,** 5037–5044 (1984).

175. Sack, R. B. & Froehlich, J. L. Antigenic similarity of heat-labile enterotoxins from diverse strains of Escherichia coli. *J. Clin. Microbiol.* **5,** 570–572 (1977).

176. Sack, D. A., Merson, M. H., Wells, J. G., Sack, R. B. & Morris, G. K. Diarrhoea associated with heat-stable enterotoxin-producing strains of Escherichia coli. *Lancet Lond. Engl.* **2,** 239–241 (1975).

177. Refined Structure of Escherichia coli Heat-labile Enterotoxin, a Close Relative of Cholera Toxin Titia K. Sixma et al., *Journal of Molecular Biology*. 1993

178. Haan, L. de & Hirst, T. R. Cholera toxin: A paradigm for multi-functional engagement of cellular mechanisms (Review). *Mol. Membr. Biol.* **21,** 77–92 (2004).

179. Del Canto, F. *et al*. Distribution of classical and nonclassical virulence genes in enterotoxigenic Escherichia coli isolates from Chilean children and tRNA gene screening for putative insertion sites for genomic islands. *J. Clin. Microbiol.* **49,** 3198–3203 (2011).

180. Roy, K. *et al*. Enterotoxigenic Escherichia coli EtpA mediates adhesion between flagella and host cells. *Nature* **457,** 594–598 (2009).

181. Sherlock, O., Vejborg, R. M. & Klemm, P. The TibA adhesin/invasin from enterotoxigenic Escherichia coli is self recognizing and induces bacterial aggregation and biofilm formation. *Infect. Immun.* **73,** 1954–1963 (2005).

182. Nicklasson, M., Sjöling, Å., von Mentzer, A., Qadri, F. & Svennerholm, A.-M. Expression of colonization factor CS5 of enterotoxigenic Escherichia coli (ETEC) is enhanced in vivo and by the bile component Na glycocholate hydrate. *PloS One* **7,** e35827 (2012).

183. Gorbach, S. L. & Tabaqchali, S. Bacteria, bile and the small bowel. *Gut* **10,** 963–972 (1969).

184. Friedman, E. S. *et al*. Microbes vs. chemistry in the origin of the anaerobic gut lumen. *Proc. Natl. Acad. Sci*. 201718635 (2018). doi:10.1073/pnas.1718635115

185. Myers, K. S. *et al*. Genome-scale Analysis of Escherichia coli FNR Reveals Complex Features of Transcription Factor Binding. *PLOS Genet*. **9,** e1003565 (2013).

186. Volbeda, A., Darnault, C., Renoux, O., Nicolet, Y. & Fontecilla-Camps, J. C. The crystal structure of the global anaerobic transcriptional regulator FNR explains its extremely fine-tuned monomer-dimer equilibrium. *Sci. Adv*. **1,** e1501086 (2015).

187. El Aidy, S., van den Bogert, B. & Kleerebezem, M. The small intestine microbiota, nutritional modulation and relevance for health. *Curr. Opin. Biotechnol*. **32,** 14–20 (2015).

188. Marteyn, B. *et al*. Modulation of *Shigella* virulence in response to available oxygen *in vivo*. *Nature* **465,** 355–358 (2010).

189. Zheng, L., Kelly, C. J. & Colgan, S. P. Physiologic hypoxia and oxygen homeostasis in the healthy intestine. A Review in the Theme: Cellular Responses to Hypoxia. *Am. J. Physiol. - Cell Physiol.* **309,** C350–C360 (2015).

190. Schwartz, C. J., Djaman, O., Imlay, J. A. & Kiley, P. J. The cysteine desulfurase, IscS, has a major role in in vivo Fe-S cluster formation in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 9009–9014 (2000).

191. Eckweiler, D., Dudek, C.-A., Hartlich, J., Brötje, D. & Jahn, D. PRODORIC2: the bacterial gene regulation database in 2018. *Nucleic Acids Res.* **46,** D320–D326 (2018).

192. Kurabayashi, K., Tanimoto, K., Tomita, H. & Hirakawa, H. Cooperative Actions of CRP-cAMP and FNR Increase the Fosfomycin Susceptibility of EnterohaemorrhagicEscherichia coli(EHEC) by Elevating the Expression of glpT and uhpT under Anaerobic Conditions. *Front. Microbiol.* **8,** 426 (2017).

193. Kumar, A., Alam, A., Rani, M., Ehtesham, N. Z. & Hasnain, S. E. Biofilms: Survival and defense strategy for pathogens. *Int. J. Med. Microbiol.* **307,** 481–489 (2017).

194. Flemming, H.-C. *et al*. Biofilms: an emergent form of bacterial life. *Nat. Rev. Microbiol.* **14,** 563–575 (2016).

195. Puorger, C., Vetsch, M., Wider, G. & Glockshuber, R. Structure, folding and stability of FimA, the main structural subunit of type 1 pili from uropathogenic Escherichia coli strains. *J. Mol. Biol.* **412,** 520–535 (2011).

196. Sheikh, A. *et al.* Highly conserved type 1 pili promote enterotoxigenic E. coli pathogen-host interactions. *PLoS Negl. Trop. Dis.* **11,** e0005586 (2017).

197. Schembri, M. A. & Klemm, P. Biofilm Formation in a Hydrodynamic Environment by Novel FimH Variants and Ramifications for Virulence. *Infect. Immun.* **69,** 1322–1328 (2001).

198. Barbieri, N. L. *et al.* FNR Regulates the Expression of Important Virulence Factors Contributing to the Pathogenicity of Avian PathogenicEscherichia coli. *Front. Cell. Infect. Microbiol.* **7,** 265 (2017).

199. Luethy, P. M. *et al.* Microbiota-Derived Short-Chain Fatty Acids Modulate Expression of Campylobacter jejuni Determinants Required for Commensalism and Virulence. *mBio* **8,** (2017).

200. Anderson, C. J., Clark, D. E., Adli, M. & Kendall, M. M. Ethanolamine Signaling Promotes Salmonella Niche Recognition and Adaptation during Infection. *PLOS Pathog.* **11,** e1005278 (2015).

201. Bertin Yolande *et al.* Enterohaemorrhagic Escherichia coli gains a competitive advantage by using ethanolamine as a nitrogen source in the bovine intestinal content. *Environ. Microbiol.* **13,** 365–377 (2010).

202. Denkel, L. A., Rhen, M. & Bange, F.-C. Biotin sulfoxide reductase contributes to oxidative stress tolerance and virulence in Salmonella enterica serovar Typhimurium. *Microbiol. Read. Engl.* **159,** 1447–1458 (2013).

203. Gupta, S. & Chowdhury, R. Bile affects production of virulence factors and motility of Vibrio cholerae. *Infect. Immun.* **65,** 1131–1134 (1997).

204. Jones, S. A. *et al.* Respiration of Escherichia coli in the Mouse Intestine. *Infect. Immun.* **75,** 4891–4899 (2007).

205. Begum, Y. A. *et al.* In Situ Analyses Directly in Diarrheal Stool Reveal Large Variations in Bacterial Load and Active Toxin Expression of Enterotoxigenic Escherichia coli and Vibrio cholerae. *mSphere* **3,** (2018).

206. Sjöling, A. *et al.* In vivo expression of the heat stable (estA) and heat labile (eltB) toxin genes of enterotoxigenic Escherichia coli (ETEC). *Microbes Infect.* **8,** 2797–2802 (2006).

207. Kaas, R. S., Leekitcharoenphon, P., Aarestrup, F. M. & Lund, O. Solving the Problem of Comparing Whole Bacterial Genomes across Different Sequencing Platforms. *PLOS ONE* **9,** e104984 (2014).

208. Jones, F. R. *et al.* New World monkey Aotus nancymae as a model for Campylobacter jejuni infection and immunity. *Infect. Immun.* **74,** 790–793 (2006).

209. Thibault, P. *et al.* Identification of the Carbohydrate Moieties and Glycosylation Motifs in Campylobacter jejuni Flagellin. *J. Biol. Chem.* **276,** 34862–34870 (2001).

210. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci.* **97,** 6640–6645 (2000).